

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:06:11 ; Search time 71 Seconds
(without alignments)
6541.772 Million cell updates/sec

Title: US-10-776-827-10

Perfect score: 8026

Sequence: 1 cggggcagcaaccaggagat.....aattaaaggatatactagctc 4460

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spool_p/US10776827/runat_21092004_133240_20767/app_query.fasta_1.4615
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1 1 120 @runat_21092004_133240_20767 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRBADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	850	10.6	153	1 Y040 HUMAN	Q15053 homo sapien
2	123.5	1.5	319	1 TCBI_RABIT	P06333 oryctolagus
3	117.5	1.5	628	1 V70K TYMV	P10357 turnip yell
4	113	1.4	628	1 V70K TYMVC	P28478 turnip yell
5	112.5	1.4	631	1 HNEFA HUMAN	P20823 ononis yell
6	112	1.4	597	1 HNEFA TYMV	P20130 ononis yell
7	108.5	1.4	1007	1 SAL2 HUMAN	Q9Y467 homo sapien
8	108	1.3	353	1 ALCL1 GORGO	P20758 gorilla gor
9	108	1.3	862	1 Z409 HUMAN	Q9up66 homo sapien
10	107.5	1.3	2063	1 NC06 HUMAN	Q14586 h nuclear r
11	107	1.3	628	1 V70K TYMVA	P20131 turnip yell
12	106.5	1.3	1004	1 SAL2 MOUSE	Q9GX96 mus musculus
13	106.5	1.3	2843	1 APC HUMAN	P25054 homo sapien
14	106	1.3	5179	1 MUC2 HUMAN	Q02817 homo sapien
15	105	1.3	1964	1 NTC4 MOUSE	P01695 mus musculus
16	103	1.3	353	1 ALCL1 HUMAN	P01876 homo sapien
17	103	1.3	917	1 SMOO HUMAN	P53814 homo sapien
18	102.5	1.3	2842	1 APC_RAT	P70478 rattus norv

c	19	102	1.3	962	1	GUNA_PSEFL	P10476 pseudomonas
	20	101.5	1.3	628	1	HNEFA_RAT	P15257 rattus norv
c	21	101.5	1.3	769	1	ITB2 BOVIN	P32592 bos taurus
	22	101.5	1.3	1185	1	DRPL HUMAN	P54259 homo sapien
	23	101	1.3	817	1	HUNB_WUSDO	Q01778 musca domes
	24	101	1.3	1959	1	AGRI_RAT	P25304 rattus norv
	25	101	1.3	2003	1	NTC4_HUMAN	Q99466 homo sapien
	26	100.5	1.3	703	1	ZMI5_HUMAN	Q9H091 homo sapien
	27	100.5	1.3	977	1	BAB1_DROME	Q9W0K7 drosophila
	28	100.5	1.3	1210	1	AF4 HUMAN	P51825 homo sapien
	29	100.5	1.3	1509	1	GSRI_HUMAN	Q9NZM4 homo sapien
	30	100	1.2	1257	1	PGCN_RAT	P55067 rattus norv
	31	99.5	1.2	648	1	FXN1_HUMAN	O15353 homo sapien
	32	99.5	1.2	660	1	YHL1_EBV	P03181 epstein-bar
	33	99.5	1.2	1229	1	P121_HUMAN	Q9Y2N3 homo sapien
	34	99	1.2	315	1	SK12_HUMAN	O15370 homo sapien
	35	99	1.2	759	1	HUNB_DROVA	O62541 drosophila
	36	99	1.2	2453	1	NCR1_MOUSE	Q60374 mus musculus
	37	98.5	1.2	2845	1	APC_MOUSE	Q61315 mus musculus
	38	98	1.2	1426	1	BCL9_HUMAN	O00512 homo sapien
	39	98	1.2	1845	1	Z236_HUMAN	Q9UL36 homo sapien
	40	97.5	1.2	448	1	EDAR_MOUSE	Q9R187 mus musculus
	41	97	1.2	1544	1	TUL4_HUMAN	Q9NRJ4 homo sapien
	42	96.5	1.2	344	1	CD2_RAT	P08921 rattus norv
	43	96.5	1.2	628	1	HNEFA_MOUSE	P22361 mus musculus
	44	96.5	1.2	693	1	CAUP_DROME	P54269 drosophila
	45	96.5	1.2	1183	1	DRPL_RAT	P54258 rattus norv

ALIGNMENTS

RESULT 1

Y040 HUMAN
ID Y040 HUMAN STANDARD; PRT; 153 AA.
AC Q15053;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0040.
GN KIAA0040.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Kagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1."
RL DNA Res. 1:27-35(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC
RX
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; D25539; BAA05022.1; -
DR EMBL; Z99715; CAB16870.1; -
KW Hypothetical protein.
SQ SEQUENCE 153 AA; 17035 MW; CAC3F8AF16821935 CRC64;

Alignment Scores:			
Pred. No.:	1..93e-66	Length:	153
Score:	850.00	Matches:	153
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	10.59%	Indels:	0
DB:	1	Gaps:	0
US-10-776-827-10 (1-4460) x Y040_HUMAN (1-153)			
Qy	816	ATGCACATATCTCCATGTCACAGAGTAACTACTCAACCAAGGAACAAACCTCAGACTAAG	875
Db	1	MetHisTyrValHisValHisArgValThrGlnProArgAsnLysProGlnThrLys	20
Qy	876	TGTCCTCCAGTGAGGGCGAGTCCCGAGGACCGCTGGAGCAATCTTGGATATCTCTGGCCA	935
Db	21	CysProSerGlyGlyGlnSerGlnGlyProArgGlyGlnPheLeuAspThrValLeuAla	40
Qy	936	GCTATGTGTCATAGCAATAGCAATGCTCTTACTGCAGACCCAGGCATGCTCCCACTGTC	995
Db	41	AlaMetCysProIleAlaMetLeuLeuThrAlaAspProGlyMetProThrCysLeu	60
Qy	996	TGGCATATCCCACTGCAAGCAACAAAGAACATTTATTCATACATCTCAATATGTTGCC	1055
Db	61	TrpHisThrProHisAlaLysHisGlyGluHisLeuSerIleHisLeuAsnMetValpro	80
Qy	1056	AAGTGTGTGCATGCAAGTAAACACACACACACAAATTCAGGTACGAGTACGTGGGC	1115
Db	81	LysCysValHisMetHisValThrHisThrHisThrAsnSerGlySerArgTyrValgly	100
Qy	1116	AAGTATATTCGCTCATCAATGCTCATTTGGCTATGTAATTTGTCAGGGAAGTACATTA	1175
Db	101	LysTyrIleLeuLeuIleLysTrpSerLeuAlaMetTyrPheValGlnGlySerThrLeu	120
Qy	1176	TCTACAGTCACAAAATGTCTCATGGGAAGCCCTTGCAGATTTCAGACACATATATACAA	1235
Db	121	SerThrValThrLysMetSerHisGlyLysAlaLeuProAspSerAspThrTyrIleGln	140
Qy	1236	TTTCTCTAACCGAAGGCCCCCATACCATCTATTCCA	1274
Db	141	PheProAsnGlnGlnGlyProHisHisThrProSerIlePro	153

```

RESULT 2
TCBI_RABIT STANDARD; PRT; 319 AA.
ID TCBI_RABIT STANDARD; PRT; 319 AA.
AC P06333;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell receptor beta chain ANA 11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=85242712; PubMed=2989826;
RA Angelillo A.L., Lamoyi E., Bernstein K.E., Mage R.G.;
RT "Identification of genes for the constant region of rabbit T-cell
RT receptor beta chains.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4498-4502(1985).
RN [2]
REVISEMENTS TO 79-83 AND 104-105.
RA Mage R.G.;
RL Submitted (APR-1987) to the PIR data bank.
RL PIR: A94052; RWRBB.
DR HSSP: P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS50835; IG LIKE; 1.
DR T-cell. Receptor; Immunoglobulin domain; Transmembrane; Glycoprotein.

```

	C REGION.					
FT DOMAIN	142	273				
FT TRANSMEM	292	319				
FT DOMAIN	314	319				CYTOPLASMIC TAIL.
FT CARBOHYD	147	147				N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	164	164				N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	254	254				N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	262	282				N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE	319 AA;	36068 MW;				AUD2C3035ED45306 CRC64;
 Alignment Scores:						
Pred. No.:	0.0165	Length:	319			
Score:	123.50	Matches:	78			
Percent Similarity:	34.22%	Conservative:	38			
Best Local Similarity:	23.01%	Mismatches:	112			
Query Match:	1.54%	Indels:	111			
DB:	1	Gaps:	15			
 US-10-776-827-10 (1-4460) x TCB1_RABIT (1-319)						
Qy	954	ATGCTCTTACTGCAGACCAGCGCATGCTCCACCTGT----	992			
Dd	24	LeuthMetThrProGluiHisAlaArGproProThrThrThraMetLeuTyAspGly	43			
Qy	993	---CTCTGGCAVACCCACATCGAAGAAGCAAAGAAATTATCCATACTCAATATG	1049			
Dd	44	PheLeuTrpSerHisThrHisArgHisSerTy-LeuLeuHisProHisThrHisVal	63			
Qy	1050	GTTCCCAGTGTCGACATGCGACGTAAACACACACACAATAATTCAGGTACAGGTAC	1109			
Dd	64	-----CysThrHisThrHisThrCys-ThrHisThrHisIleHis-	76			
Qy	1110	GTGGCGAAGTATATCTGCTCATCAATGGTCTATGGCTATGTACTTGTGCGAGGAAGT	1169			
Dd	77	----AlaserThrHisValCysIleHisThrHisThrPhetrHisIleuCys----	93			
Qy	1170	ACATTATCTACAGTCACAAAAATGTCTATGGAAAAGCCTTGGCAGATTTCAGACATAT	1229			
Dd	93	eHisThrLeuThrHisAlaLeuThrLeuThrCysAlaProThrArgThrTyrrAlaHis--	112			
Qy	1230	ATACAAATTCGTACACGACGAGCCCCATACACCATCTATTCCATAAACCACTCAGTT	1289			
Dd	113	-----ThrArgAlaProThrHisValHisProHisLysProArgProAr	127			
Qy	1290	ACAGATGCATGCTTTCTCTATTCTTAACCTCACACATAAACTTTTACTGGAGTACTCAT	1349			
Dd	127	gGlueuSerAlalaLeuleuLeuProThrProLeuHisPhetrProgluAspLeuAlaAs	147			
Qy	1350	ATTGGACATTCACGAA-----CCTGCTACAGTC-----	1378			
Dd	147	nValserAlaPro-GlnValValValPheAspProSerGluAlaGluIleAasnLysThrG	167			
Qy	1379	-----CCACCCCTGTGTGCTTCATACAGAC-----	1405			
Dd	167	InLysAlaThrLeuValCysLeuAlaLysaspPheTyProAspHisValGluLeuSerT	187			
Qy	1406	-----ACACCAAGTTTTCTGTGCTCTGACCCCTCACCTACCTGTGCCAA	1445			
Dd	187	xPtpValAsnGlyLysGluValHisAsnGlyValSerThrAspProGlnProTyrrLysG	207			
Qy	1446	GATGTTAAAGTGTGATGGTTCAAAATTCATTGAAGCTCTTTCTT-----	1492			
Dd	207	InAspProLysSerAspHisIserLysTyCysLeuSerSerArgLeuArgValserAlaa	227			
Qy	1493	--GTAACCTCATGACAAAAGTCGCTCCTCAT-----TGCCACTGAGAGGTGTTAATGTGA	1544			
Dd	227	IaPheTrpPhiSASNPro----ArgasNHISPhetrGysGlnValGlnPheHeGlyLeut	246			
Qy	1545	TCCAAGACTCTCTGTGAAACATTACCCCGCAACACTCAGCAAGAGTGCCTTCTCCA	1604			
Dd	246	hrAspAspGluTrpThrTyrrAenSerSerLysProIleThr-----	260			
Qy	1605	ACCAAGACAAGAGACTCTTGGTGGTGACTGTAGAAATTTATGGAAGCCCACTCATTTA	1664			


```

Db 401 LeuLeuProAsnProProAlaAala---LeuProSerIleAlaTyrThrSerSerArgGly 419
QY 3765 AGG 3767
Db 420 Lys 420
RESULT 4
W70K_TYMCV
ID W70K_TYMCV STANDARD; PRT; 628 AA.
AC P28478;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE 01-DEC-1992 (Rel. 24, Last annotation update)
DE 69 kDa protein.
OS Turnip yellow mosaic virus (isolate TYMC).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;
OC Tymovirus.
OX NCBI_TaxID=31751;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119261; PubMed=1731998;
RA Dreher T.W., Branson K.L.;
RT "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a
RL cDNA-based clone with verified infectivity.";
RL Plant Mol. Biol. 18:403-406(1992).
CC -!- FUNCTION: Not known.
CC -!- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16378; CAA34414.1; -.
CC F01; S19150; S19150.
CC InterPro: IPR004935; Tymo 45 70kDa.
CC Pfam: PF03251; Tymo_45kd_70kd; 1.
CC SEQUENCE 628 AA; 69057 MW; 0DF0C64E8ECBC6DC CRC64;
Alignment Scores:
Pred. No.: 0.168 Length: 628
Score: 113.00 Matches: 103
Percent Similarity: 31.53% Conservative: 37
Best Local Similarity: 23.20% Mismatches: 139
Query Match: 1.41% Indels: 165
DB: 1 Gaps: 22
US-10-776-827-10 (1-4460) x V70K_TYMCV (1-628)
QY 2748 ATCCCTTGTCACATAGAGTGTGTTCATGGGCTC-----CTGGCTGTGCCCTTC 2798
Db 40 LeuProMetValHisSerGluGlyThrSerThrProThrGlnLeuLeuArgHisProAsn 59
QY 2799 ACATAACAGATAACTTGGCCATCTGCTGCTGCACCAACCCAGGATGTGGAGAGATCTCC 2858
Db 60 IleTrpPheGlyAsnLeuProProProProArgArgProGlnAspAsnArgAspPheSer 79
QY 2859 CCACAACTGCCACTGCTC----- 2876
Db 80 ProLeuHisProLeuValPheProGlyHisHisSerGlnLeuArgHisValHisGluThr 99
QY 2877 -----ACCAGACAGAGTGCCTTCTCTCTCCACCTCTCAGTCCCTAGATGGATG 2930
Db 100 GlnGlnValGlnGlnThrCysProGlyLysLeuLysLeuSer----- 113
QY 2931 GCTGGGGAGAGGTGAGGCTGACAGCTGAGAGCTAGTGTGATGATGATGATGAGGAGGCG 2990
Db 114 -----GlyAla 115
QY 2991 GATCACCAGGATCCGGGACCACATACAGTAACATGCT-----TTCCATGGCACTGCTTG 3044

```

```

Db 116 GluGluLeuProProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPr 135
QY 3045 CTGG-----TTTGAATTAAAGACAGCAGTCAGTT----- 3072
Db 135 oSerArgPheProHisPheHisAlaArgArgProAspValLeuProSerValProAs 155
QY 3073 -----GTCATTGCCATGACAAAGCCCTCTATCT----- 3099
Db 155 pHisGlyProValLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerAl 175
QY 3100 -----CCAGGCACAATGTCCTGCTGCTCTCTAAT----- 3129
Db 175 aThrArgGlyProSerPheArgProIleLeuLeuProLysValValHisValHisAspAs 195
QY 3130 -----CCAATGGACTGCTCTCACCCAGGGATGAACACCCAGAACTCACT----- 3177
Db 195 pProProHisSerSerLeuArgProArgLysArgSerArgSerArgGlnLeuGlnProThrVa 215
QY 3178 -----TCTCAGTCACCTTCCACAGCCGATGAC 3203
Db 215 lArgArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe 235
QY 3204 TCAGAGAGCCAAACCCAGATGGGCTCTCTTTC----- 3240
Db 235 rAspAspProGlyIleLeuLeuGlyProArgProLeuAlaProAsnSerThrArgAspProPr 255
QY 3241 -----CCATCAGACGCTCCCTGCACAAAC-----CTTCTCTGGCG 3275
Db 255 oProArgProIleThrProGlyProSerAsnThrHisGlyLeuArgProLeuSerValIe 275
QY 3276 TAACTAGAGGAGTCCAGTGCAGGATAGGCCCTTAAACGTTTGTAAATAAACAGGTGCA 3335
Db 275 uProArgAlaSerProArgArgGly-----LeuLeuProAsnProArgArgHi 291
QY 3336 TGAAGAGGAGCTTAAGGCCATTTGTGATATCCACTCTCTCTTCTTCCACTT-----CCTTC 3389
Db 291 sArgThrSerThrGlyHisIleProProThrThrThrSerArgProThrGlyProProSe 311
QY 3390 TCATCTTTTCTCCATGTTTATGCTTCTGATTCCTCTCTGCTGCTGCACAGACCAG 3449
Db 311 rArgLeuGlnArgProVal-----HisLeuTyrGlnSerSe 323
QY 3450 CCCAGCCCTTTATCTCTCTCCATTTTCACTCTCCAGCTC----- 3492
Db 323 rPro-----HisThrProAsnPheArgProSerSerIleArgLysAspAlaLeuLe 340
QY 3493 -----TGCTCCTGAACTGCCACTGCCACTGCCAACCCATG 3521
Db 340 uGlnThrGlyProArgLeuGlyHisLeuGluCys-----LeuGlyGlnProAl 356
QY 3522 GGACCTCAGGACCCAGAGACTGCTTGACTCATCTGGGAGGGTAAAGTTTCACGGGGGACAAA 3581
Db 356 aAsnLeuArgThr-----SerGluArgSerProPr 366
QY 3582 AAAATGATTCCTAAAGAAGAGGCTTCTAGACAGACAGACAGCTCCAGAAAGACATCCCT 3641
Db 366 o-----ThrLysArgArgLeuProArgSerSerGlu---ProAsnArg--LeuPro 381
QY 3642 AGGCTTGGACTTCTGACAGCTTTAGCCAGCTCCGACCGCAGCCAGCAGGAGGAGGCTTTC 3701
Db 382 LysPro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisArgSerTyr 399
QY 3702 CCCATTGCTCTTCTCCCATTTGCTCAATGGATTCATGTTTCTTTTCTTTTGGGGGGAGCA 3761
Db 400 ProPheLeuProAsnProProAlaAla---LeuProSerIleAlaTyrThrSerSerArg 418
QY 3762 GGGAGG 3767
Db 419 GlyLys 420
RESULT 5
HNFA_HUMAN

```


ID HNFA HUMAN STANDARD; PRT; 631 AA.
 AC P20823; Q99861;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
 DE transcription factor LF-B1) (LFB1) (Transcription factor-1)
 GN TCF1 OR HNF1A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=91184801; PubMed=1707031;
 RA Guenot J.-L., Galcheva-Gargova Z., Mattei M.-G., Simon-Chazottes D.,
 RA Guenet J.-L., Cereghini S., Yaniv M.;
 RT "Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal
 RT localization of its gene in man and mouse.";
 RL Genomics 8:155-164(1990).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94038905; PubMed=7900999;
 RA Bach I., Yaniv M.;
 RT "More potent transcriptional activators or a transdominant inhibitor
 RT of the HNF1 homeoprotein family are generated by alternative RNA
 RT processing.";
 RL EMBO J. 12:4229-4242(1993).
 [3]
 RP SEQUENCE FROM N.A., AND VARIANT MODY3 LEU-447.
 RX MEDLINE=97100943; PubMed=8945470;
 RA Yamagata K., Oda N., Kaisaki P.J., Menzel S., Furuta H.,
 RA Vaxillaire M., Southern L., Cox R.D., Lathrop G.M., Boriraj V.V.,
 RA Chen X., Cox N.J., Oda Y., Yano H., le Beau M.M., Yamada S.,
 RA Nishigori H., Takeda J., Fajans S.S., Hattersley A.T., Iwasaki N.,
 RA Hansen T., Pedersen O., Polonsky K.S., Turner R.C., Velho G.,
 RA Chevre J.-C., Froguel P., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-
 RT onset diabetes of the young (MODY3).";
 RL Nature 384:455-458(1996).
 [4]
 RP VARIANTS MODY3 ARG-107; TRP-131; MET-260 AND HIS-272.
 RX MEDLINE=97309269; PubMed=9166684;
 RA Glucksmann M.A., Lehto M., Tayber O., Scotti S., Berkemeter L.,
 RA Pulido J.C., Wu Y., Nir W.-J., Fang L., Markel P., Munnelly K.D.,
 RA Goranson J., Orho M., Young B.M., Whitacre J.L., McMenimen C.,
 RA Wantman M., Tuomi T., Warram J., Forablon C.M., Carlsson M.,
 RA Rosenzweig J., Kennedy G., Duyk G.M., Krolewski A.S., Groop L.C.,
 RA Thomas J.D.;
 RT "Novel mutations and a mutational hotspot in the MODY3 gene.";
 RL Diabetes 46:1081-1086(1997).
 [5]
 RP VARIANTS MODY3 HIS-12; GLN-131; GLN-205 AND CYS-263, AND VARIANT NIDDM
 ASP-191.
 RX MEDLINE=97431548; PubMed=9287053;
 RA Iwasaki N., Oda N., Ogata M., Hara M., Hinokio Y., Oda Y.,
 RA Yamada S., Nishigori H., Ohgawara H., Omori Y., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha/MODY3 gene in
 RT Japanese subjects with early- and late-onset NIDDM.";
 RL Diabetes 46:1504-1508(1997).
 [6]
 RP VARIANT NIDDM MET-254, AND VARIANTS LEU-27 AND ASN-487.
 RX MEDLINE=97431550; PubMed=9287055;
 RA Yamada S., Nishigori H., Onda H., Takahashi K., Kitano N.,
 RA Morikawa A., Takeuchi T., Takeda J.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene (MODY3) are
 RT not a major cause of late-onset NIDDM in Japanese subjects.";
 RL Diabetes 46:1512-1513(1997).
 [7]
 RP VARIANTS IDDM HIS-272 AND GLY-583.
 RX MEDLINE=97458988; PubMed=9313763;

RA Yamada S., Nishigori H., Onda H., Utsugi T., Yanagawa T., Maruyama T.,
 RA Onigata K., Nagashima K., Nagai R., Morikawa A., Takeuchi T.,
 RA Takeda J.;
 RT "Identification of mutations in the hepatocyte nuclear factor (HNF)-1-
 RT alpha gene in Japanese subjects with IDDM.";
 RL Diabetes 46:1643-1647(1997).
 [8]
 RP VARIANTS MODY3, AND VARIANT ATYPICAL DIABETES SER-574.
 RX MEDLINE=98052398; PubMed=9392505;
 RA Boutin P., Chevre J.-C., Hani E.H., Gomis R., Pardini V.C.,
 RA Guillausseau P.-J., Vaxillaire M., Velho G., Froguel P.;
 RT "An automated fluorescent single-strand conformation polymorphism
 RT technique for screening mutations in the hepatocyte nuclear
 RT factor-1alpha gene (maturity-onset diabetes of the young).";
 RL Diabetes 46:2108-2109(1997).
 [9]
 RP VARIANTS MODY3 GLN-131; GLN-229; GLY-241 AND HIS-272.
 RX MEDLINE=97184312; PubMed=9032114;
 RA Kaisaki P.J., Menzel S., Lindner T., Oda N., Rjasanowski I., Sahn J.,
 RA Meincke G., Schulze J., Schmechel H., Petzold C., Ledermann H.M.,
 RA Sachse G., Boriraj V.V., Menzel R., Kerner W., Turner R.C.,
 RA Yamagata K., Bell G.I.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene in MODY and
 RT early-onset NIDDM: evidence for a mutational hotspot in exon 4.";
 RL Diabetes 46:528-535(1997).
 [10]
 RP VARIANTS MODY3 THR-129; TRP-131; TRP-159; LEU-519 AND ILE-620.
 RX MEDLINE=97230240; PubMed=9075818;
 RA Frayling T.M., Bulman M.P., Ellard S., Appleton M., Dronsfield M.J.,
 RA Mackie A.D., Baird J.D., Kaisaki P.J., Yamagata K., Bell G.I.,
 RA Bain S.C., Hattersley A.T.;
 RT "Mutations in the hepatocyte nuclear factor-1alpha gene are a common
 RT cause of maturity-onset diabetes of the young in the U.K.";
 RL Diabetes 46:720-725(1997).
 [11]
 RP VARIANTS MODY3 ASN-128; TYR-143 AND LEU-447.
 RX MEDLINE=97230241; PubMed=9075819;
 RA Hansen T., Eiberg H., Rouard M., Vaxillaire M., Moeller A.M.,
 RA Rasmussen S.K., Fridberg M., Urhammer S.A., Holst J.J., Almind K.,
 RA Schwald S.M., Hansen L., Bell G.I., Pedersen O.;
 RT "Novel MODY3 mutations in the hepatocyte nuclear factor-1alpha gene:
 RT evidence for a hyperexcitability of pancreatic beta-cells to
 RT intravenous secretagogues in a glucose-tolerant carrier of a P447L
 RT mutation.";
 RL Diabetes 46:726-730(1997).
 [12]
 RP VARIANTS LEU-27; VAL-98 AND ASN-487.
 RX MEDLINE=97278987; PubMed=9133564;
 RA Urhammer S.A., Fridberg M., Hansen T., Rasmussen S.K., Moeller A.M.,
 RA Clausen J.O., Pedersen O.;
 RT "A prevalent amino acid polymorphism at codon 98 in the hepatocyte
 RT nuclear factor-1alpha gene is associated with reduced serum C-peptide
 RT and insulin responses to an oral glucose challenge.";
 RL Diabetes 46:912-916(1997).
 [13]
 RP VARIANT NIDDM GLN-583, AND VARIANTS LEU-27; VAL-98 AND ASN-487.
 RX MEDLINE=97266232; PubMed=9112026;
 RA Urhammer S.A., Rasmussen S.K., Kaisaki P.J., Oda N., Yamagata K.,
 RA Moeller A.M., Fridberg M., Hansen L., Hansen T., Bell G.I.,
 RA Pedersen O.;
 RT "Genetic variation in the hepatocyte nuclear factor-1 alpha gene in
 RT Danish Caucasians with late-onset NIDDM.";
 RL Diabetologia 40:473-475(1997).
 [14]
 RP VARIANTS MODY3 CYS-122; PHE-142 AND GLN-159.
 RX MEDLINE=97253393; PubMed=9097962;
 RA Vaxillaire M., Rouard M., Yamagata K., Oda N., Kaisaki P.J.,
 RA Boriraj V.V., Chevre J.-C., Boccio V., Cox R.D., Lathrop G.M.,
 RA Dussoix P., Philippe J., Timsit J., Charpentier G., Velho G.,
 RA Bell G.I., Froguel P.;
 RT "Identification of nine novel mutations in the hepatocyte nuclear
 RT factor 1 alpha gene associated with maturity-onset diabetes of the
 RT young (MODY3).";

RL	Hum. Mol. Genet. 6:583-586 (1997).	345	GAGGAATAAAATTACAAGTGGTCAAGTGGTCAATCTTTGAGCTCAGGACCTGGCTGTGTAAG	404
RN	[15]	Db		
RP	VARIANTS LEU-27; ASN-487 AND ARG-514.	228	GluArg---GluThrLeuValGluGluCys-----AsnArgAlaGluCysle	242
RP	MEDLINE=98265912; PubMed=9604876;	Qy		
RA	Behn P.S., Wasson J., Chayen S., Smolovitch I., Thomas J.D.,	405	CCGAGAGGGTCTCTGGCCCTAAATCTAGCCAGCACCATGAGAGAAATCACTGCTTCTT	464
RA	Glaser B., Permut M.A.;	Db		
RT	"Hepatocyte nuclear factor lalpa coding mutations are an uncommon	243	GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu	261
RT	contributor to early-onset type 2 diabetes in Ashkenazi Jews.;"	Qy		
RL	Diabetes 47:967-969 (1998).	465	CAGCTCTATCTGGGACACCATCTTGCACCAACACCAAGAGGCATCTTACAACACCATCTG	524
RN	[16]	Db		
RP	VARIANTS MODY3 ASP-31; TRP-159; THR-161; TRP-200 AND TRP-271.	262	ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu	281
RX	MEDLINE=98425814; PubMed=9754819;	Qy		
RA	Chevre J.-C., Hani E.H., Boutin P., Vaxillaire M., Blanche H.,	525	-----CCTGGGAGTCTCTCTGGGCTCTGGCTCTGGT	557
RA	Vionnet N., Pardini V.C., Timsit J., Langer E., Charpentier G.,	282	AlaMetAspThrTyrSerGlyProProGlyProGlyProGlyProAla---LeuPro	300
RA	Beckers D., Maes M., Bellanne-Chantelot C., Velho G., Froguel P.;	Qy		
RT	"Mutation screening in 18 Caucasian families suggest the existence of	558	GATCATCACACTCTCTTCTCATCTGTGTCATTTGCTGGAGCCACACAGGAGGGG	617
RT	other MODY genes.;"	Db		
RL	Diabetologia 41:1017-1023 (1998).	301	AlaHisSerSerPro---GlyLeuProProAlaLeuSerProSerLysValHisGly	319
RN	[17]	Qy		
RP	VARIANTS IDDM LYS-48 AND GLY-241.	618	-----CCAGCAGCCAGAGAGAAAGAGAGAGAGAGAGAGAGATGA	662
RX	MEDLINE=99083111; PubMed=9867222;	Db		
RA	Moeller A.M., Dalgaard L.T., Pociot F., Nerup J., Hansen T.,	320	ValArg**GlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly---	338
RA	Pedersen O.;	Qy		
RT	"Mutations in the hepatocyte nuclear factor-lalpa gene in Caucasian	663	AGAAGACCTCTGGATCTCTGCTCAACC-----CAAGCTTCTCCAGATGAGAGAG	713
RT	families originally classified as having type I diabetes.;"	Db		
RL	Diabetologia 41:1528-1531 (1998).	339	---GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlu	357
RN	[18]	Qy		
RP	VARIANTS MODY3 ARG-537 AND LYS-619.	714	ACATCATCTGCTGTATTAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	773
RX	MEDLINE=98289379; PubMed=9626139;	Db		
RA	Elbein S.C., Teng K., Yount P., Scroggin E.;	358	Pro---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly-----	374
RA	"Linkage and molecular scanning analyses of MODY3/hepatocyte nuclear	Qy		
RT	factor-1 alpha gene in typical familial type 2 diabetes: evidence for	774	TCCTTCTGACCCACACACACACATTTTCAGGAACCCCTGAAATATATGCACTATGTC	833
RT	novel mutations in exons 8 and 10.;"	Db		
RL	J. Clin. Endocrinol. Metab. 83:2059-2065 (1998).	375	-----GlyProLeuProProValSerThrLeuThrAlaLeu	386
RN	[19]	Qy		
RP	VARIANTS LEU-27 AND ASN-487.	834	CACAGA-----GTAATCTACTCAACCAAGGAGCAACACCTCAGACT	872
RX	MEDLINE=98284534; PubMed=9621514;	Db		
RA	Nishigori H., Yamada S., Kohama T., Uteugi T., Shimizu H.,	387	HisSerLeuGluGlnThrSerProGlyLeuAsnGlnGlnProGlnAsnLeuLeuMetAla	406
RA	Takeuchi T., Takeda J.;	Qy		
RT	"Mutations in the hepatocyte nuclear factor-1 alpha gene 'MODY3' are	873	AGTGTCTCCAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	905
RT	not a major cause of early-onset non-insulin-dependent 'type 2'	407	SerLeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThr	425
RT	diabetes mellitus in Japanese.;"	Db		
RL	J. Hum. Genet. 43:107-110 (1998).	905	-----	905
RN	[20]	Qy		
RP	VARIANTS MODY3 HIS-12; ASN-158; GLN-159 AND CYS-203.	426	PheThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSer	445
RX	MEDLINE=99176457; PubMed=10078571;	Qy		
RA	Yamada S., Tomura H., Nishigori H., Shio K., Mabe H., Iwatani N.,	906	-----CCTGGACAAATCTCTGGATCTCTCTGGCAGCT-----	938
RA	Takumi T., Kito Y., Moriya N., Muroya K., Ogata T., Onigata K.,	446	ValProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSer	465
RA	MoriKawa A., Inoue I., Takeda J.;	Qy		
RT	"Identification of mutations in the hepatocyte nuclear factor-lalpa	939	-----ATGTGTCCAATAGCAATGCTCTTACT	965
RT	gene in Japanese subjects with early-onset NIDDM and functional	466	GlnProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThr	485
RT	analysis of the mutant proteins.;"	Qy		
RL	Diabetes 48:645-648 (1999).	965	GCAGACCCAGGATGCTCTCCACCTGCTCTGGCATACCCCATGCA-----AAGCAC	1019
RN	[21]	Db		
RP	VARIANTS MODY3 ASP-31; TRP-159; THR-161; TRP-200 AND TRP-271.	486	GlnSerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHis	505
RX	MEDLINE=98425814; PubMed=9754819;	Qy		
RA	Chevre J.-C., Hani E.H., Boutin P., Vaxillaire M., Blanche H.,	1020	AAAGAACATTTATCC-----ATACATCTCAATATGGTTCCTCCAGTGTGTGCACATGCAC	1073
RA	Vionnet N., Pardini V.C., Timsit J., Langer E., Charpentier G.,	Db		
RA	Beckers D., Maes M., Bellanne-Chantelot C., Velho G., Froguel P.;	506	LysProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeu	523
RT	"Mutation screening in 18 Caucasian families suggest the existence of	1074	GTAACACACACACACACAAATTCAGGTAGCAGGTACGTGGGCAAGTATATTCTGCTCATC	1133
RT	other MODY genes.;"	Qy		
RL	Diabetologia 41:1017-1023 (1998).	524	IleThrAspThr-----	527
RN	[17]	Db		
RP	VARIANTS IDDM LYS-48 AND GLY-241.	1134	AAATGGTCATTGGCTATGTACTTTGTCAGGGAAAGTACATTATCTACAGTCAACAAATG	1193
RX	MEDLINE=99083111; PubMed=9867222;	Qy		
RA	Moeller A.M., Dalgaard L.T., Pociot F., Nerup J., Hansen T.,	528	-----ThrAsnLeuSerAlaLeuSerLeu	536
RA	Pedersen O.;	Db		
RT	"Mutations in the hepatocyte nuclear factor-lalpa gene in Caucasian	1194	TCTCATGGGAAGCCTTCCAGATTCACACACATATACAAATTTCTTCTTCTTCTTCT	1253
RT	families originally classified as having type I diabetes.;"	Qy		
RL	Diabetologia 41:1528-1531 (1998).	225	GAGATCCCTGACACCAAGCAGAGTAACTGCGAGGAGGCTCCCAAGTGGCTG	284
RN	[18]	Qy		
RP	VARIANTS MODY3 ARG-537 AND LYS-619.	193	GlnLeuPro---ThrLysLysGlyArgArgAsnArgPheLysTyrGlyProAlaSerGln	211
RX	MEDLINE=98289379; PubMed=9626139;	Qy		
RA	Elbein S.C., Teng K., Yount P., Scroggin E.;	285	GCCATTTTGTGCTGT	344
RA	"Linkage and molecular scanning analyses of MODY3/hepatocyte nuclear	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RT	factor-1 alpha gene in typical familial type 2 diabetes: evidence for	Db		
RT	novel mutations in exons 8 and 10.;"	Qy		
RL	J. Clin. Endocrinol. Metab. 83:2059-2065 (1998).	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RN	[19]	Db		
RP	VARIANTS LEU-27 AND ASN-487.	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RX	MEDLINE=98284534; PubMed=9621514;	Qy		
RA	Nishigori H., Yamada S., Kohama T., Uteugi T., Shimizu H.,	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RA	Takeuchi T., Takeda J.;	Qy		
RT	"Mutations in the hepatocyte nuclear factor-1 alpha gene 'MODY3' are	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RT	not a major cause of early-onset non-insulin-dependent 'type 2'	Qy		
RT	diabetes mellitus in Japanese.;"	Db		
RL	J. Hum. Genet. 43:107-110 (1998).	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RN	[20]	Qy		
RP	VARIANTS MODY3 HIS-12; ASN-158; GLN-159 AND CYS-203.	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RX	MEDLINE=99176457; PubMed=10078571;	Qy		
RA	Yamada S., Tomura H., Nishigori H., Shio K., Mabe H., Iwatani N.,	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RA	Takumi T., Kito Y., Moriya N., Muroya K., Ogata T., Onigata K.,	Qy		
RA	MoriKawa A., Inoue I., Takeda J.;	Db		
RT	"Identification of mutations in the hepatocyte nuclear factor-lalpa	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RT	gene in Japanese subjects with early-onset NIDDM and functional	Qy		
RT	analysis of the mutant proteins.;"	Db		
RL	Diabetes 48:645-648 (1999).	212	GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227
RN	[21]	Qy		

Alignment Scores:		
Pred. No.:	0.186	Length: 631
Score:	112.50	Matches: 103
Percent Similarity:	31.55%	Conservative: 44
Best Local Similarity:	22.10%	Mismatches: 180
Query Match:	1.40%	Indels: 139
DB:	1	Gaps: 23

US-10-776-827-10 (1-4460) x HNFA_HUMAN (1-631)		
Qy	225 GAGATCCCTGACACCAAGCAGAGTAACTGCGAGGAGGCTCCCAAGCTGGCTG	284
Db	193 GlnLeuPro---ThrLysLysGlyArgArgAsnArgPheLysTyrGlyProAlaSerGln	211
Qy	285 GCCATTTTGTGCTGT	344
Db	212 GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu	227

QY 1230 ATACAATTTCTTACACAGGAGGCCCCATACACCATCTTATCCATTAACCACTCAGGTT 1289
 Db 110 -----ProSerProSerThrProProThrProSerProPro----- 121
 QY 1290 ACAGATGATGCTTTCTTATTTCTTAACTACACATAAACTTTTACTGGAAGTACTCAT 1349
 Db 122 -----CysCysHisPro---ArgLeuSerLeuHisArgProAlaLeuGluAspLeu 138
 QY 1350 ATTGGACATTCAGCAACCTG----- 1370
 Db 139 LeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyLeuArgAspAlaSerGlyVal 158
 QY 1371 -----CTACAGTCCCCACCTTGTGTGTC 1394
 Db 159 ThrPheThrTrpThrProSerSerGlyLeuValGluGlyProProGluArgAsp 178
 QY 1395 TTGATACAGACACCAAGCTTTCTGTGCTCTGACCCCTCCTCCTGCGCAAGATTTAA 1454
 Db 179 LeuGlySerGlyCysTyzSerValSerValLeuPro---GlyCysAlaGluProTrpAs 197
 QY 1455 AGTGTGATGTTCAAAATTCATGAAGCTCTTTCTTGTAACTCATGACAAAGTCCGTC 1514
 Db 197 n-----HisGlyLysThrPheThrCysThrAlaAlaLysProGlu 211
 QY 1515 CTCATTGCCACTGAGAGGTTTAATGTGATCCA-----AGACCTCTCTGTGCA 1562
 Db 211 rLysThrProLeuThrAlaThrLeuSerLysSerGlyAsnMetPheArgProGluValHi 231
 QY 1563 AACATTACCCCGCAACCACTCAG 1587
 Db 231 sLeuLeuProProProSerGluGlu 239
 RESULT 9
 Z409 HUMAN STANDARD; PRT; 862 AA.
 AC Q9UPU6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein 409.
 GN ZNF409 OR KIAA1056.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205 (1999).
 CC -!- FUNCTION: May function as a transcription factor.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: Contains 6 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. Usage by and for commercial
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; AB028979; BAA83008.1; --
 DR HSSP; P15822; 3ZNF.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00096; zF-C2H2; 5.
 DR SMART; SM00355; Znf_C2H2; 6.

DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; FALSE NEG.
 KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT ZN_FING 237 259 C2H2-TYPE 1.
 FT ZN_FING 453 476 C2H2-TYPE 2.
 FT ZN_FING 508 532 C2H2-TYPE 3.
 FT ZN_FING 569 593 C2H2-TYPE 4.
 FT ZN_FING 758 782 C2H2-TYPE 5.
 FT ZN_FING 821 845 C2H2-TYPE 6.
 FT DOMAIN 605 710 PRO-RICH.
 SQ SEQUENCE 862 AA; EB6BF9A0B939217 CRC64;
 Alignment Scores:
 Pred. No.: 0.506 Length: 862
 Score: 108.00 Matches: 97
 Percent Similarity: 28.25% Conservative: 53
 Best Local Similarity: 18.27% Mismatches: 158
 Query Match: 1.35% Indels: 223
 DB: 1 Gaps: 22
 US-10-776-827-10 (1-4460) x Z409_HUMAN (1-862)
 QY 26 GGGCTTGAGAGAGCCCTTCCGCGACCGAAGATTGTT----- 64
 Db 223 GlyProMetGlyAsnSerGlyGlyAsnHisValAlaValPheTrpLeuCysLeuLeuCys 242
 QY 65 -----CCCCATTTGGAGATGAAGAACTCGAGACTCAAAGCAGCTGA 106
 Db 243 ArgLeuGlyPheSerLysProGlnAlaPheMetAspHisThrGlnSerHisGlyValLys 262
 QY 107 GTGACCTTCCCAAGGACACACACTGAACCTGGCGGTGATCAGATCTGAATGCACAGGGC 166
 Db 263 LeuThr---ProAlaGlnTyrGlnGlyLeuSerGlySerProAlaValLeuGlnGluGly 281
 QY 167 -----GGGTGTTTCAGGATTTGTTTACTAGCTTGAACGTGACCTCCAGGAAGCAGTTCT 220
 Db 282 AspGluGlyCysLysAlaLeuLeuSerPheLeuGluProLysLeuProAlaArgProSer 301
 QY 221 GCGCGAGATCCCTCGACACGCAAGCAAGTAAGTAACTGGGAAGGA----- 266
 Db 302 -SerAspIleProLeuAsnSerSerThrValAsnMetGluAlaAsnValAlaGlnTh 321
 QY 267 -----GGCTCCCA-----AGCTGGTGGCCATTGCTCTGTGTGTGGAGTGCT 313
 Db 321 rGluAspGlyProProGluAlaGluValGlnAlaLeuLeuLeuLeuAspGluValMe 341
 QY 314 GTCAGTGGCATGCCAAACCCA-----AAGCTGGA 343
 Db 341 tAlaLeuSerProProSerProProThrAlaThrTrpAspProSerProThrGlnAlaLys 361
 QY 344 AGAGGAATRAATTACAAAGTGTCAAGTTGTCATCTCTTTCAGCTCTATCTGGGACACCATCT 395
 Db 361 sGluSerProValAlaAlaGlyGlu-----AlaGlyProAspTrpPh 375
 QY 396 -----GCTTGAAGCCGAGAGGGTTCTTCTGGCCCTAA 427
 Db 375 eProGluGlyGlnGluAspGlyGlyLeuCysProProLeuAsnGlnSerSerProTh 395
 QY 428 TCTAGCCCAACCACTGAGAGATCATGCTCTTCTCAGCTCTATCTGGGACACCATCT 487
 Db 395 r---SerLysGluGlyGly-----ThrLeuProAlaProValGlySerProGl 410
 QY 488 TGACCAAAACCAAGAGGATCTTACACACCATCTTCTCTGGGAGTCTCTCTGGGCTGC 547
 Db 410 uAsp-----ProSerAspProProGlnProTy 419
 QY 548 CACTCTTTGTGATCATCATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 607
 Db 419 rArgLeuAlaAspAspTyrThrPro----- 427
 QY 608 GCAAGAGGGGCCAGCAGCCAG 667

Db 428 -----AlaProAlaala----- 431
QY 668 ACCTCTGGATCTCTGCTCAACCAAGCTTCTCCAGATGGAGAGACACCACTACTGCTG 727
Db 431 ----- 431
QY 728 TTTAGTTAGGAGAGAGAGAGAGTGTTCCTTTCTGGGGCTTAAGCTCTCTCTGACCACA 787
Db 432 -----PheGlnGly----- 434
QY 788 CACAGACATTTAGGAGACCTCCGAATGATGCTATGTCATGTCACAGAGTAACATAC 847
Db 435 -LeuSerLeuSerSerHisMetSerLeuLeuHis----- 445
QY 848 TCAACCAAGGAACAAACCTCAGACT--AAGTCTCCAGTGGAGGCGCAGTCCCGAGGACC 904
Db 446 ----SerArgAsnSerCysLysThrLeuLysCysProLys----- 457
QY 905 ACGTGGACAATTTCTGGATCTCTCTTGGCAGCTATGTCCTCAATAGCAATGCTCCTTAC 964
Db 457 ----- 457
QY 965 TGCAGACCCAGGATGCTCCACCTGCTCTGCGCATACCCACATGCAAGCAACAGAA 1024
Db 459 -----CysAsnTrp-----HisTyLysTyrglnGln 466
QY 1025 ACATTTATCCATACATCTCAATATGTTCCCAAGTGTGTCACATGCACGTAAACACAC 1084
Db 466 nThrLeuAsp-----ValHisMetGlnGlyHisPr 477
QY 1085 ACACACAATTCAGGTAGCAGGTAC----- 1109
Db 477 oGluSerAsnSerHisCysSerTyrcysSerAlaGlyGlyAlaHisProArgLeuAlaR 497
QY 1110 -----GTGGCGAAGTATATTC 1126
Db 497 gGlyGluSerTyrcysGlyTyrcysProTyrcysArgCysAspValCysAsnTyrcerth 517
QY 1127 GCTCATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1186
Db 517 rThrThrLysGlyAsnLeuSerHisMetGlnSerAspLysHisLeuAlaAsnLeuGln 537
QY 1187 AAAATGTCTCATGGGAAGCCCTTCCAGATTCAGACACATATATCAATTTCTCTAAACA 1246
Db 537 nGlyPheGlnAlaGlyProGlyGlyGlnGlySerProThrGluAlaSerLeuProProSe 557
QY 1247 GCAAGGCCCATACACCATCTATTTCCATAACCACTCAGGTACAGATGCTGCTTTCC 1306
Db 557 rAlaGlyAspLysGluProLysThrLysSerSerTrpGln---CysLysValCysSerTy 576
QY 1307 TATTCTTAACCTACACATAAACTTTTACTGGAAGTACTCATATTTGACATTCAGCAA 1366
Db 576 rGluThrAsnLeuSerArgAsnLeuArgGlnHisMetThrSerGluLysHisMetGlnAs 596
QY 1367 CCGTCTA-----CAGTCCCGCCAGCTTGT 1388
Db 596 nValLeuMetLeuHisGlnGlyLeuProLeu 606
RESULT 10
NC06_HUMAN STANDARD; PRT; 2063 AA.
ID Q14686; Q9NT29; Q9UW74; Q9UK86;
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)
DE (Cancer-amplified transcriptional coactivator ASC-2) (Activating
DE signal integrator-2) (ASC-2) (Peroxisome proliferator-activated
DE receptor-interacting protein) (PPAR-interacting protein) (PRIP)
DE (Nuclear receptor-activating protein, 250 kDa) (Nuclear receptor
DE coactivator RAP250) (NRC RAP250) (Thyroid hormone receptor-binding
DE protein).
GN NCOA6 OR AIB3 OR RAP250 OR TRBP OR KIAA0181.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthalia; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH CREBBP; NCOA1; GTF2A; TBP;
RP RXRA; ESR1; RARA AND THRA.
RX MEDLINE=20036574; PubMed=10567404;
RA Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
RA Jung Y.-K., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
RA Junn B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
RT "A nuclear factor ASC-2, as a cancer-amplified transcriptional
RT coactivator essential for ligand-dependent transactivation by nuclear
RT receptors in vivo.";
RL J. Biol. Chem. 274:34283-34293(1999).
RN [2]
RP SEQUENCE FROM N.A., HOMODIMERIZATION, AND INTERACTION WITH CREBBP;
RP RXRA; ESR1; NR3C1; RARA; VDR AND THRA.
RX MEDLINE=20325329; PubMed=10866662;
RA Mahajan M.A., Samuels H.H.;
RT "A new family of nuclear receptor coregulators that integrates nuclear
RT receptor signaling through CBP.";
RL Mol. Cell. Biol. 20:5048-5063(2000).
RN [3]
RP SEQUENCE FROM N.A., AND INTERACTION WITH PPARG; ESR1; ESR2 AND
RP THR.
RC TISSUE=Testis;
RX MEDLINE=20148724; PubMed=10681503;
RA Cairns F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT "Cloning and characterization of RAP250, a nuclear receptor
RT coactivator.";
RL J. Biol. Chem. 275:5308-5317(2000).
RN [4]
RP SEQUENCE FROM N.A., PHOSPHORYLATION BY PRKDC, AND INTERACTION WITH
RP THR; RAR; EP300 AND CRSP3.
RC TISSUE=Lymphocytes;
RX MEDLINE=20283976; PubMed=10823961;
RA Ko L., Cardona G.R., Chin W.W.;
RT "Thyroid hormone receptor-binding protein, an LXXLL motif-containing
RT protein, functions as a general coactivator.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6212-6217(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RL analysis of cDNA clones from human cell line KG-1.";
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dharm P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor J.,
RA Levenshao M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Sruce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RA "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [7]
RP INTERACTION WITH NCOA6IP.
RX MEDLINE=21417756; PubMed=11517327;
RA Zhu Y.-J., Qi C., Cao W.-Q., Yeldandi A.V., Rao M.S., Reddy J.K.;
RT "Cloning and characterization of PIMT, a protein with a
RT methyltransferase domain, which interacts with and enhances nuclear
RT receptor coactivator PRIP function.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:10380-10385(2001).
RN [8]
RP INTERACTION WITH RBM14.
RX MEDLINE=21423995; PubMed=11443112;
RA Iwasaki T., Chin W.W., Ko L.;
RT "Identification and characterization of RRM-containing coactivator
RT activator (CoAA) as TRBP-interacting protein, and its splice variant
RT as a coactivator modulator (CoAM)." ;
RL J. Biol. Chem. 276:33375-33383(2001).
RN [9]
RP INTERACTION WITH HRMTLL1.
RX MEDLINE=22151129; PubMed=12039952;
RA Qi C., Chang J., Zhu Y., Yeldandi A.V., Rao S.M., Zhu Y.-J.;
RT "Identification of protein arginine methyltransferase 2 as a
RT coactivator for estrogen receptor alpha." ;
RL J. Biol. Chem. 277:28624-28630(2002).
RN [10]
RP INTERACTION WITH MLL3 AND THE ASCOM COMPLEX.
RX TISSUE=Cervical carcinoma;
RX MEDLINE=22371496; PubMed=12482968;
RA Goo Y.-H., Sohn Y.-C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,
RA Azores D.O., Meltzer P.S., Sun P.-G., Song E.J., Lee K.-J., Lee Y.C.,
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins." ;
RL Mol. Cell. Biol. 23:140-149(2003).
RN [11]
RP MUTAGENESIS OF 883-THR-GLU-894, AND PHOSPHORYLATION.
RX MEDLINE=21635582; PubMed=11773444;
RA Ko L., Cardona G.R., Iwasaki T., Bramlett K.S., Burris T.P.,
RA Chin W.W.;
RT "Ser-884 adjacent to the LXXLL motif of coactivator TRBP defines
RT selectivity for ERs and TRs." ;
RL Mol. Endocrinol. 16:128-140(2002).
CC -|- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
CC receptors and stimulates the transcriptional activities in a
CC hormone-dependent fashion. Coactivates expression in an agonist-
CC and AP2-dependent manner. Involved in the coactivation of
CC different nuclear receptors, such as for steroids (GR and ERs),
CC retinoids (RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR)
CC and prostanoids (PPARs). Probably functions as a general
CC coactivator, rather than just a nuclear receptor coactivator. May
CC also be involved in the coactivation of the NF-kappa-B pathway.
CC May coactivate expression via a remodeling of chromatin and its
CC interaction with histone acetyltransferase proteins.
CC -|- SUBUNIT: Monomer and homodimer. Interacts with RNPC2 (By
CC similarity). Interacts in vitro with the basal transcription
CC factors GTF2A and TBP, suggesting an autonomous transactivation
CC function. Interacts with NCOA1, CRSP3, RBM14, the histone
CC acetyltransferases EP300 and CREBBP, and with the
CC methyltransferases NCOA6IP and HRMTLL1/PRMT2. Belongs to the
CC ASC-2/NCOA6 complex (ASCOM), which contains ASC-2/NCOA6, the
CC retinoblastoma-binding protein RBQ-3/ RBBP5, alpha- and beta-
CC tubulins, the trithorax group proteins MLL2 and MLL3, and
CC ASH2/ASCL2.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain,
CC prostate, testis and ovary; weakly expressed in lung, thymus and
CC

small intestine.
-|- DOMAIN: Contains two Leu-Xaa-Xaa-Leu-Leu (LXXLL) motifs. Only
motif 1 is essential for the association with nuclear receptors,
while adjacent Ser-884 displays selectivity for nuclear receptors.
-|- PFM: Phosphorylated by PRKDC.
-|- PFM: Phosphorylation on Ser-884 leads to a strong decrease in
binding to ESRI and ESR2.
-|- MISCELLANEOUS: Frequently amplified or overexpressed in colon,
breast and lung cancers.
-|- CAUTION: Ref.1 (AAFI6403) sequence differs from that shown due to
a frameshift in position 88.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
EMBL; AF177388; AAF13595.1; -;
DR EMBL; AF208227; AAF16403.1; ALT_FRAME.
DR EMBL; AF245115; AAF78480.1; -;
DR EMBL; AF128458; AAF37003.1; -;
DR EMBL; AF171667; AAF71829.1; -;
DR EMBL; D80003; BAA11498.2; ALT_INIT.
DR EMBL; AL109824; CAB92721.1; -;
DR Genew; HGNC:15936; NCOA6.
DR MIM; 605299; -;
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; TAS.
DR GO; GO:0003682; F:chromatin binding; ISS.
DR GO; GO:0003031; F:retinoid X receptor binding; TAS.
DR GO; GO:0046965; F:retinoid hormone receptor binding; IDA.
DR GO; GO:0046966; F:thyroid hormone receptor binding; IDA.
DR GO; GO:0003713; F:transcription co-activator activity; IDA.
DR GO; GO:0016563; F:transcriptional activator activity; TAS.
DR GO; GO:0007420; P:brain development; ISS.
DR GO; GO:0001701; P:embryonic development (sensu Mammalia); ISS.
DR GO; GO:0007507; P:heart development; ISS.
DR GO; GO:0030099; P:myeloid blood cell differentiation; IDA.
Alignment Scores:
Pred. No.: 0.722 Length: 2063
Score: 107.50 Matches: 97
Percent Similarity: 34.70% Conservative: 47
Best Local Similarity: 23.37% Mismatches: 130
Query Match: 1.34% Indels: 141
DB: Gaps: 23

US-10-776-827-10 (1-4460) x NCO6_HUMAN (1-2063)
QY 2683 CCTAAGCACTCAGGCTCCACATC-----ATCAAC 2712
Db 1043 ProLysSerValArgLeuProValSerGlnAsnValHisProProArgGlyProLeuAsn 1062
QY 2713 CCCTTTTACCACAGAGA-----AAGAGCACTCTGGTTCTCTATCCCTTGTGCATAGAG 2766
Db 1063 ProAspSerGlnArgMetProMetGlnGlnSerGlySerVal---ProValMetVal--- 1080
QY 2767 AGTTTGTATGGGGCCCTCTGGCTGTGGCTTCCATACATACAGATAACTTGCATCTGCCT 2826
Db 1081 SerLeuGlnGlyProAlaSerValProProSerProAspLysGlnArgMetProMetPro 1100
QY 2827 GCACCAACACCA-----GGGATGTGGAAGACATCTCCCAAC 2865
Db 1101 ValAsnThrProLeuGlySerAsnSerArgLysMetValTyrGlnGluSerProGlnAsn 1120
QY 2866 TGGCACTGCTCACCA-----GGACACAGCTGCCCTTCT---GTCTCCACCTCTCAGTCC 2916
Db 1121 ProSerSerProLeuAlaGluMetAlaSerLeuProGluAlaSerGlySerGluAla 1140
QY 2917 CCCTAGATGATGGCTGGGAGAGGTGGAGCTGACAGCTGAGACGTAGTGTGATAT 2976

```

Db 1141 Pro-----1141
QY 2977 GATCTAGGAGCGGATCAGCGGATCCGGACCATCAAGTAAACATGTTTCATGGCA 3036
Db 1142 -----SerValProGlyGlyProAsnAsnMetProSerHisVal 1154
QY 3037 ACTGCTTGCTCGTTGAATTAAGACAGCAGTCAGTTGTCATTCGCAAGCCCTTA 3096
Db 1155 val-----LeuProGlnAsnGlnLeuMet-----MetThrGlyPro--- 1166
QY 3097 TCTCCAGGCACAATGTCCTGCTGTCTCTTAATCAATGCAATGCTCTCACCCCGGGA 3156
Db 1167 -----LysProGlyProSerProLeuSerAlaThrGlnGly 1178
QY 3157 TGAACACCCAGAACTCACTTCTCAGTCACTTCCACAG-----3195
Db 1179 AlaThrProGlnGlnProProValAsnSerHisGlyHisHisPhePro 1198
QY 3196 -----CCGATGACTCAGAGAGGCAACCCAGAAATGGGCTCTCTTTTC 3240
Db 1199 AsnValAlaAlaProThrGlnThrSerArgProLysThrProAsnArgAlaSerProArg 1218
QY 3241 CCC---ATCACAGACTCCCTGACAACTTCTCTGGCGTAACCTAGAGAGTCCAGTGCA 3297
Db 1219 ProTyrTyrProGlnThrProAsnAsnArg-ProProSerThr-GluProSerGluLeu 1238
QY 3298 GGATAGGCCCTAAACGTTTGTGTTAAATAACAGGTGCATGAAGGAGCTAAGGCCATTG 3357
Db 1238 erLeuSerProGluArg-----1243
QY 3358 TTGATATCCACTCTCTTCTTCCACTTCTCTCATCTTTTCTCCATGTTTATGCTTC 3417
Db 1244 -----LeuAsnAlaSerIleAlaGlyLeuPhePro-----1253
QY 3418 TCTGATTCCTCTCTGCTGCACAGACCCAGCCCTTATTTCTCTCCATTTT- 3476
Db 1254 -----ProGlnIleAsnIleProLeuProProArgProAsnLeuAsnArgGlyPheA 1271
QY 3477 -----CACTCTTCCAGCTCTGCTCCCTG-----AACTGCC 3507
Db 1271 spGlnGlnGlyLeuAsnProThrThrLeuLeuAlaIleGlyGlnAlaProSerAsnLeu 1291
QY 3508 ACTGGCAACCCATGGGA-----CTCAGGACCCAGAGCTG-----CTTGACT 3549
Db 1291 hrMetAsnProSerAsnPheAlaThrProGlnThrHisLysLeuAspSerValValAla 1311
QY 3550 CATCTGGGAGGGAAGTTACGGGGGACAAATAATGATTCTTAAGAAGAGGCTCTCT 3609
Db 1311 snSerGlyLysGlnSerAsnSerGlyAlaThrLysArgAlaSerProSerAsnSerArgA 1331
QY 3610 AGACGACACAGGC---TCCAGAAAGACATCCCTTAGCGCTGGAGTTCTGAGCAGCTTTA 3666
Db 1331 rgSerSerProGlySerSerArgLysThrThrProSerProGlyArgGlnAsn-----S 1349
QY 3667 GCCAGGCTCCGAC-----GGCAGCCAGAGG-----3692
Db 1349 erLysAlaProLysLeuThrLeuAlaSerGlnThrAsnAlaLeuLeuGlnAsnValG 1369
QY 3693 -----AGCGCTTCCCATGCTCTCTTCCCATGCT 3725
Db 1369 luLeuProArgAsnValLeuValSerProThrProLeuAla 1382

RESULT 11
V70K_TYMVA
ID V70K_TYMVA STANDARD; PRT; 628 AA.
AC P20131;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE 69 kDa protein.
OS Turnip yellow mosaic virus (Australian isolate).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymoviridae;

```

```

OC Tymovirus.
OX NCBI_TaxID=12155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90021184; PubMed=2800335;
RA Keese P., Mackenzie A., Gibbs A.;
RT "Nucleotide sequence of the genome of an Australian isolate of turnip
RL yellow mosaic tymovirus.";
RL Virology 172:536-546(1989).
CC -1- SIMILARITY: Not known.
CC -1- SIMILARITY: TO 65 TO 70 kDa PROTEIN FROM OTHER TYMOVIRUSES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; J04373; AAA46591.1; -.
DR PIR; JQ0110; JQ0110.
DR InterPro; IPR004935; Tymo_45_70kDa.
DR Pfam; PF03251; Tymo_45kd_70kd; 1.
SQ SEQUENCE 628 AA; 68740 MW; 67CD342A09161D96 CRC64;

```

Alignment Scores:

```

Pred. No.: 0.565 Length: 628
Score: 107.00 Matches: 96
Percent Similarity: 30.57% Conservative: 37
Best Local Similarity: 22.07% Mismatches: 155
Query Match: 1.33% Indels: 147
DB: 1 Gaps: 17

```

US-10-776-827-10 (1-4460) x V70K_TYMVA (1-628)

```

QY 2748 ATCCCTTGTGCATAGAGAGTTTGTCTATGGGCGCT-----CTGGCTGTGCCCTTC 2798
Db 40 LeuProMetValHisSerGluGlyThrSerAlaProThrGlnLeuLeuArgHisProAsn 59
QY 2799 ACATAACAGATAACTTTCCTCCATCTCCCTGCCAACCAACCCAGGAGTGTGAAGACATCTCC 2858
Db 60 IleTrpPheGlyAsnIleProProProArgArgProGlnAspAsnArgAspPheSer 79
QY 2859 CCACAACTGCCACTG---CTCACGAGGACAAAGCTGCCCTCTCTGTCTCCACCTCTCAGTC 2915
Db 80 ProLeuHisProLeuValPhe-ProGlyHisHisSer-----91
QY 2916 CCCCTAGAAATGGATGGCTGGGGAGAGGTGGAGGTGACAGCTGAGACGTAGTGTGAGATA 2975
Db 92 -----GlnLeuArg-----94
QY 2976 TGATCTAGGAGGCGGATCACCGGGATCCGGGACCATACACAGTAACTGTTTTCATGGC 3035
Db 95 -----HisValHisGluThrGlnGlnValG1 103
QY 3036 AACTGCTTGTCTGTTGAATTA-----3057
Db 103 nGlnThrCysProGlyGluLeuLysLeuSerGlyLeuGluLeuProProAlaProG1 123
QY 3058 -AGACAGAGTCACTTGTCTATTTGTCATGACAGCCCTCTATCT---CCAGGCACAATGTC 3113
Db 123 nArgGlnHisSerLeuProLeuHisIleThrArgProSerArgPheProHisHisPheHi 143
QY 3114 CCTGCTGTCTTAATCAATGAGTCTCTCTACCCAGGAGTGAACACCCAGAACT 3173
Db 143 sAlaArgArgProAspValLeuProSerLeuProAspHisGly-----ProVa 159
QY 3174 CACTTCTCAGTCACTTCCACAG---CCGATGACTCAGAAGAGCCCAACCCAGATGGGC 3230
Db 159 lLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerThrThrArgGlyPr 179
QY 3231 CTCTCTTTTCCCATCACA-----GACTCCCTCGACACCT 3266

```


RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
RP AXIN.

RP MEDLINE=20271867; PubMed=10811618;
RA Spink K.E., Polakis P., Weis W.I.;
RA "Structural basis of the axin-adenomatous polyposis coli
RT interaction.";
RL EMBO J. 19:2270-2279 (2000).

RP REVIEW ON VARIANTS.

RP MEDLINE=94154728; PubMed=8111410;
RA Nagase H., Nakamura Y.;
RA "Mutations of the APC (adenomatous polyposis coli) gene.";
RL Hum. Mutat. 2:425-434 (1993).

RP VARIANTS FAP.

RP MEDLINE=91335211; PubMed=1651563;
RA Nishisho I., Nakamura Y., Miyoshi Y., Ando H., Horii A.,
RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.;
RA "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
RT patients.";
RL Science 253:665-669 (1991).

RP VARIANTS FAP.

RP MEDLINE=93265030; PubMed=1338904;
RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
RA Miyoshi Y., Mori T., Nakamura Y.;
RA "Somatic mutations of the APC gene in colorectal tumors: mutation
RT cluster region in the APC gene.";
RL Hum. Mol. Genet. 1:229-233 (1992).

RP VARIANTS FAP.

RP MEDLINE=93244793; PubMed=1338691;
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
RA Nakamura Y., Horii A., Aoki T., Petersen G.M.,
RA Vogelstein B., Maher E., Ogawa M., Utsunomiya J.,
RA Baba S., Nakamura Y.;
RA "Screening for germ-line mutations in familial adenomatous polyposis
RT patients: 61 new patients and a summary of 150 unrelated patients.";
RL Hum. Mutat. 1:467-473 (1992).

RP VARIANTS FAP TRP-99.

RP MEDLINE=95134544; PubMed=7833149;
RA Dobbie Z., Spycher M., Huerliman R., Ammann R., Roth J.,
RA Mueller A., Mueller H., Scott R.J.;
RA "Mutation analysis of the first 14 exons of the adenomatous
RT polyposis coli (APC) gene.";
RL Eur. J. Cancer 30A:1709-1713 (1994).

RP VARIANTS FAP GLY-722.

RP MEDLINE=95135430; PubMed=7833931;
RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.;
RA "Four novel mutations of the APC (adenomatous polyposis coli) gene in
RT FAP patients.";
RL Hum. Mol. Genet. 3:1687-1688 (1994).

RP ERRATUM.

RA Stella A., Montero M., Resta N., Marchese C., Susca F., Gentile M.,
RA Romio L., Pilia S., Prete F., Marenzi C., Guanti G.;
RL Hum. Mol. Genet. 3:1918-1918 (1994).

RP VARIANT FAP ILE-171.
RX MEDLINE=97144176; PubMed=8990002;
RA van der Luijt R.B., Meera Khan P., Vassen H.F.A., Tops C.M.J.,
RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
RA Plug R.J., Griffioen G., Rodde R.;
RA "Molecular analysis of the APC gene in 105 Dutch kindreds with
RT familial adenomatous polyposis: 67 germline mutations identified by
RT DGGE, PTT, and southern analysis.";
RL Hum. Mutat. 9:7-16 (1997).

RP VARIANTS COLORECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.

RP MEDLINE=98080146; PubMed=9419979; Kikuchi-Yanoshita R., Tanaka K.,
RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
RA Fukutome A., Tomiyama J., Chuganji Y., Momoi M., Utsunomiya J.;
RA "Drastic genetic instability of tumors and normal tissues in Turcot
RT syndrome.";
RL Oncogene 15:2877-2881 (1997).

RP VARIANT LYS-1307.

RP MEDLINE=98400248; PubMed=9731522;
RA Redston M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Satagopan J.,
RA Wong N., Yang D., Nafa D., Abrahamson J., Ozelik H.,
RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.;
RA "The APC 11307K allele and breast cancer risk.";
RL Nat. Genet. 20:13-14 (1998).

RP VARIANTS LYS-1307 AND GLN-1317.

RP TISSUE=Peripheral blood;
RX MEDLINE=98393712; PubMed=9724771;
RA Frayling I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;
RA "The APC variants 11307K and E1317Q are associated with colorectal
RT tumors, but not always with a family history.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727 (1998).

RP VARIANT LYS-1307.

RP MEDLINE=98400259; PubMed=9731533;
RA Woodage T., King S.M., Wacholder S., Hartge P., Struwing J.P.,
RA McAdams M., Laken S.J., Tucker M.A., Brody L.C.;
RA "The APC 11307K allele and cancer risk in a community-based study of
RT Ashkenazi Jews.";
RL Nat. Genet. 20:62-65 (1998).

RP Alignment Scores:

Pred. No.:	0.97	Length:	2843
Score:	106.50	Matches:	75
Percent Similarity:	35.41%	Conservative:	33
Best Local Similarity:	24.59%	Mismatches:	105
Query Match:	1.33%	Indels:	92
DB:	1	Gaps:	14

US-10-776-827-10 (1-4460) x APC_HUMAN (1-2843)

QY 2797 TCACATACAGATAACTTCCCATCTGCTGCACCAACCCAGGATGTGGAGACATCT 2856

Db 2225 SerArgGlyArgThrMetIleHisIleProGlyValArgAsnSerSerSerThrSer 2244

QY 2857 CCCACAACTGCCACTGCTCCACGAGCAAGCTGCCCTCTCTCCACTCTCAGTCC 2916

Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260

QY 2917 CCCTAGATCGATGCTGGGAGAGTGGAGCTGACAGCTGACAGCTAGTGTGACATAT 2976

Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280

QY 2977 GATCTA-----GGAGGCGGATC----- 2994

Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlySerSerLysAlaProSer 2300

[illegible]

RA Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 RL the NOTCH gene family (NOTCH4).";
 RN Oncogene 14:1893-1890 (1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung, and Testis;
 RX MEDLINE=96281668; PubMed=8681805;
 RA Uttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
 RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
 RL cell-specific mammalian Notch gene.";
 RN Development 122:2251-2259 (1996).
 [4]
 RP SEQUENCE FROM N.A.
 RC Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Region of the mouse major histocompatibility locus class III
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases."
 RN [5]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/int3
 RL gene in a mouse mammary tumor: Generation of truncated Notch4/int3
 RN mRNAs by retroviral splicing events.";
 J. Virol. 73:5166-5171 (1999).
 [6]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uttendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 RL Notch4 in embryonic endothelium.";
 RN Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648 (2001).
 [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 OF VAL-1463.
 RX MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeder E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RL proteolysis.";
 RN J. Biol. Chem. 276:40268-40273 (2001).
 [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE=21374376; PubMed=11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RL among mammalian Notch family members.";
 CC Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
 CC Jaggel, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -! SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC -! SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytic processing NICD is translocated to the nucleus.
 CC -! TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -! DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 dpc.
 CC -! PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following

ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXTR). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane.
 -! PTM: Phosphorylated.
 -! DISEASE: Loss of the extracellular domain causes constitutive
 activation of the Notch protein, which leads to hyperproliferation
 of glandular epithelial tissues and development of mammary
 carcinomas.
 -! SIMILARITY: Belongs to the NOTCH family.
 -! SIMILARITY: Contains 29 EGF-like domains.
 -! SIMILARITY: Contains 3 Lin/Notch repeats.
 -! SIMILARITY: Contains 5 ANK repeats.
 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

EMBL; M80456; AAB38377.1; -
 EMBL; U43691; AAC52630.1; -
 EMBL; U43691; AAC52631.1; -
 EMBL; AF030001; AAB82004.1; -
 EMBL; AB016771; BAA32281.1; ALT SEQ.
 EMBL; AB016772; BAA32283.1; ALT INIT.
 EMBL; AB016773; BAA32284.1; ALT INIT.
 EMBL; AB016774; BAA32285.1; -
 PIR; A38072; TMVMT3.
 PIR; T09059; T09059.
 HSP; P08709; 1BF9.
 MGD; MGI:107471; Notch4.
 InterPro; IPR002110; ANK.
 InterPro; IPR000152; ASX_hydroxyl_S.
 InterPro; IPR000742; EGF 2.
 InterPro; IPR001881; EGF CA.
 InterPro; IPR001438; EGF II.
 InterPro; IPR006209; EGF_Like.
 InterPro; IPR002049; Laminin_EGF.
 InterPro; IPR008297; Notch.
 InterPro; IPR008900; Notch_dom.
 Pfam; PF00023; ank; 6.
 Pfam; PF00008; EGF; 27.
 Pfam; PF00066; notch; 2.
 PIRSF; PIRSF002279; Notch; 1.
 PRINTS; PR00011; EGFBLD.
 PRINTS; PR01452; NOTCH.
 SMART; SM00248; ANK; 6.
 SMART; SM00179; EGF CA; 11.
 SMART; SM00004; NL; 2.
 PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 PROSITE; PS50088; ANK_REPEAT; 5.
 PROSITE; PS00010; ASX_HYDROXYL; 11.
 PROSITE; PS00022; EGF_1; 28.
 PROSITE; PS01186; EGF_2; 21.
 PROSITE; PS50026; EGF_3; 27.
 PROSITE; PS01187; EGF_CA; 9.
 PROSITE; PS00026; EGF_3; 27.
 REceptor; Transcription regulation; Activator; Differentiation;
 Developmental protein; Repeat; ANK repeat; EGF-like domain;
 Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 SIGNAL 1 20
 CHAIN 21 1964
 CHAIN 1411 1964
 CHAIN 1428 1964
 CHAIN 1463 1964
 CHAIN 21 1443
 TRANSMEM 1444 1464
 DOMAIN 1465 1964

FT	DOMAIN	21	60	EGF-LIKE 1.	106	862	-----SerPheGlnCysLeuCysLeuG	869
FT	DOMAIN	61	112	EGF-LIKE 2.	110	1609	TTGCTTGGAGAAAGGCACCTTTGCTGAGTGGTTTCGGGGGTAATGTTTCACAGAGAGTC	1550
FT	DOMAIN	115	152	EGF-LIKE 3.	118	869	lnGlyThrThrGlyAlaLeuCysAspPheProLeuSerCysGlnMetAlaMetSerG	889
FT	DOMAIN	153	189	EGF-LIKE 4.	126	1549	TTGATCACATTAACACACCTCTCAGTGGCAATAGGACGACGACTTTGTCATGAG-----	1497
FT	DOMAIN	191	229	EGF-LIKE 5.	144	889	lnGlylleGluileSerGlyLeuCysGlnAsn---GlyGlyLeuCysleAspThrGlyS	908
FT	DOMAIN	231	271	EGF-LIKE 6.	152	1496	-----TTACAGAAAGAGCTTTCATGAATTTGA	1466
FT	DOMAIN	273	309	EGF-LIKE 7.	160	908	erSerTyrPheCysArgCysProGlyPheGlnGlyLeuCysGlnAspMetA	928
FT	DOMAIN	311	350	EGF-LIKE 8.	168	1465	AC-----CATCACATTTAAACATCTTGCCACAGGTGGGGTCT	1427
FT	DOMAIN	352	388	EGF-LIKE 9.	176	928	snProCysGluProAsnProCysHisGlySerThrCysValProSerGlyT	948
FT	DOMAIN	389	427	EGF-LIKE 10.	184	1426	AGAGCACAGAAACTTGTGTCTGTATCAAGACACAAAGGTGGGGACTGTAGCAGG	1367
FT	DOMAIN	429	470	EGF-LIKE 11.	192	948	yr-----Val-CysGlnCysAlaProGlyTyrGluGlyGlnAsnCysSerlys	963
FT	DOMAIN	472	508	EGF-LIKE 12.	200	1366	TTGCTGGAA---TGTCCAATTATGAGTACTTCCAGTAAAGTTTATGTGTAGAGTTAGAA	1310
FT	DOMAIN	508	546	EGF-LIKE 13.	208	964	ValLeuGluAlaCysGlnSerGlnProCysHisAsnHisGlyThrCysThrSerArgPro	983
FT	DOMAIN	548	584	EGF-LIKE 14.	216	1309	ATAGGAAAGCATGTCATCTGTAACTGAGTGGTTTATGGAATAGATAGTGTATGGGGGCT	1250
FT	DOMAIN	586	622	EGF-LIKE 15.	224	984	GlyGlyPheHisCysAlaCysPro-----	992
FT	DOMAIN	623	656	EGF-LIKE 16.	232	1249	TGCTGGTTAGGAAATTGTATATATGTCTGTAATCTGCAAGGCTTCCCATGAGCATT	1190
FT	DOMAIN	658	686	EGF-LIKE 17.	240	993	GlyPheValGly-----	996
FT	DOMAIN	688	724	EGF-LIKE 18.	248	1189	TTTGACTGCTGATAGTAATGTAATCTCCCTGCACAAAGTACATAGCAATGACCATTTGATG	1130
FT	DOMAIN	726	762	EGF-LIKE 19.	256	997	-----LeuArgCysGluGlyAspVal-----AspGluCysLeu	1007
FT	DOMAIN	764	800	EGF-LIKE 20.	264	1129	AGCAGAAATATCTTCCACG-----TACCTGCTACCTGAATTTGGT	1088
FT	DOMAIN	803	839	EGF-LIKE 21.	272	1008	AspArgProCysHisProSerGlyThrAlaAlaCysHisSerLeuAlaAsnAlaPheTyr	1027
FT	DOMAIN	841	877	EGF-LIKE 22.	280	1087	TGTGTGTGTGTACGTGCATGTGCACACCTTGGGACCATATTGAGATGT-----	1037
FT	DOMAIN	878	924	EGF-LIKE 23.	288	1028	Cys-----GlnCysLeuProGlyHisThrGlyGln-----ArgCysGluValGlu	1042
FT	DOMAIN	926	962	EGF-LIKE 24.	296	1036	ATGGATAAATGTTCTTTG-----TGCTTTGTCATGTGGGGTATGCCAGACAGAGTGGGA	983
FT	DOMAIN	964	1000	EGF-LIKE 25.	304	1043	MetAspLeuCysGlnSerGlnProCysSerAsnGlyGlySerCysGluileThrThrGly	1062
FT	DOMAIN	1002	1040	EGF-LIKE 26.	312	982	GGCATGCTGGGTCTGCAGTAAAGAGCATTTGCTATTGGAACATAGCTGCCAAGACAGTA	923
FT	DOMAIN	1042	1081	EGF-LIKE 27.	320	1063	ProProProGly-----	1067
FT	DOMAIN	1083	1122	EGF-LIKE 28.	328	922	TCCAAGAAATGTCACAGTGTCTCCCTGGAGCTGCCCTCCACTGGGACACTTAGTC-----	869
FT	DOMAIN	1126	1167	EGF-LIKE 29.	336	1068	ThrCysHisCysProlysGly---PheGluGlyProThrCysSerHisLysAlaLeuSer	1086
FT	REPEAT	1168	1208	LN/NOTCH 1.	344	868	-----TGAGGTTCCTTCTTGGTTGAGTAGTAGTCTCTCTGCTGG	833
FT	REPEAT	1209	1242	LN/NOTCH 2.	352	1087	CysGlylleHisHisCysHisAsnGlyGlyLeu-----	1097
FT	REPEAT	1243	1282	LN/NOTCH 3.	360	832	ACATGACATAGTGCATATTATTTCA-----GGGGTTCTGAAATGTCGTGTGTGGT	782
						1098	-----CysLeuProSerProlysProGlySerProProLeuCysAlaCys---	1112
						781	CAGAGAGGCTTAGCCCCAGAAAGGAAACACCTCTGCTTCTCTGCTTAACCTAAACAGCA	722
						1113	-----LeuSerGlyPheGlyGlyProAspCysLeuThrProProAla	1126
						721	GTGATGTCTCTTCTCCATCTGGAGAGCTTGGGTTTGAGCAGAGATCCAGAGGTCTTCTT	662
						1127	ProProGlyCysGlyProProSerProCys-----	1137

Alignment Scores:
 Pred. No.: 1.18 Length: 1964
 Score: 105.00 Matches: 171
 Percent Similarity: 27.84% Conservative: 57
 Best Local Similarity: 20.88% Mismatches: 278
 Query Match: 1.31% Indels: 3.3
 DB: 1 Gaps: 44

US-10-776-827-10 (1-4460) x NTC4_MOUSE (1-1964)

QY	2026	ACAAGCTTACCTTAAACACATCATATCCACAGTACACAGTACCTAGTTAGTACC	1967
Db	751	ThrCysLeuPro-----SerHisThrGlyArgHisCysGlnThrAlaValAsp---	766
QY	1966	TACTGTGTCAAGGCACCTTGCCTGAGCTTCAATATAATAATAATAAATCTTTATAGC	1907
Db	767	HisCysValSerAlaSerCysLeuAsnGlyGlyThrCysValAsnLysProGlyThrPhe	786
QY	1906	TTTCATCTTA-----CTTTCATTTAACTCTGTGTTACTCTCT-----	1873
Db	787	PheCysLeuCysAlaThrGlyPheGlnGlyLeu-HisCysGluGluLysThrAsnProse	806
QY	1872	-----GCATCCACCTGTGGTACCGTGCACATCACAA-----	1837
Db	806	rCysAlaAspSerProCysArgAsnLysAlaThrCysGlnAspThrProArgGlyAlaAr	826
QY	1836	-----CCTTATGACATCACTTCCAGTGTTCCTTCTGTGAGACTGCATGAT	1790
Db	826	gCysLeuCysSerProGlyTyrThrGlySerSer-----CysGlnThrLeuileAs	843
QY	1789	GATGTGTACTTCTTCCCTGCTGGAAACCTTTGGTTATATAGGAGTACCTAGCTAGGT	1730
Db	843	pLeuCysAlaArgLysProCys-----ProHisThrAla-ArgC	856
QY	1729	GTAGCATCTCTCACCTTTCCATCTCTGTAACATTTGCACAGGTACACAGTTGCAGTCCAC	1670
Db	856	ysLeuGlnSerGlyPro-----	861
QY	1669	TGACATAAATGAGTGGGCTTCCATAATTTCTAGCAGTCCACCAAGAGCTCTTTGTC	1610

```
QY 661 CAT-----CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG--- 626
Db 1138 HisAsnGlyThrCysThrGluThrProGlyLeuGlyAsnProGlyPheGlnCysThrCys 1157
QY 625 -----GCTGCTGCCCCCTCTTGC-----CTGGTGGGCTCC 596
Db 1158 ProProAspSerProGlyProArgCysGlnArgProGlyAlaSerGlyCysGluGlyArg 1177
QY 595 AGCAGCAATGGCAACAGATGAACAGAGTGTGTATGATCAACAGAGTGCAGGCCAGGA 536
Db 1178 GlyGlyAspGlyThrCysAspAlaGly-----CysSerGlyProGly 1191
QY 535 GGACTCCACAGGACATGG--TGTTGTAGATGCTTCTTGGTGTGGTCAAGATCG--- 482
Db 1192 -----GlyAspTrpAspGlyGlyAspCysSerLeuGlyValProAspProTrpLys 1208
QY 481 ---TGTCACAGATPAGAGTCAAGAGGACACTGATCTCTCCATGGTGTGGCTAGATTAA 425
Db 1209 GlyCysProHisSer-----GlnCysTrpLeuLeuPheArgAsp 1222
QY 424 GGGCCAGAGAACCT----- 410
Db 1223 GlyArgCysHisProGlnCysAspSerGluGluCysLeuPheAspGlyTyrAspCysGlu 1242
QY 409 CTCGGCTTACAAGCAGGCTCTGAGCTCAAAAGGATGCAACCTTGACCACTTGTATTTAT 350
Db 1243 IleProLeuThrCysIleProIleThrAspGln-----Tyr 1254
QY 349 TCCTCTTCCAGCTTTGGTTTGGCATGCCACT----- 317
Db 1255 CysArgAspHisPheHisAsnGlyHisCysGluLysGlyCysAsnAsnAlaGluCysGly 1274
QY 316 ---GACAGACCTCCACACACAGCAGCAAAATGGCCAGCAGCTTGGGAGCTCTCTTCC 260
Db 1275 TrpAspGlyGlyAspCysArgProGluGlyGluAspSerGlu---GlyArgProSerLeu 1293
QY 259 AGCTTACTTCTTGTCTTGGT-----TGTCAGGGATCTCGGCCAGAACTGCT 212
Db 1294 AlaLeuLeuValValLeuArgProProAlaLeuAspGlnGlnLeuAlaLeuAlaArg 1313
QY 211 TTCCTGGAGGTCACTTCAACCTAGTAAACAATCGCTGAACACCGCCCTGTGCATTTCAG 152
Db 1314 ValLeuSerLeuThrLeuArgVal-----GlyLeuTrpValArg 1326
QY 151 ATCTGTATCACCGCCAGTTCAGTGTGTCTTGGGAAGTCACTCAGCTGCTTTGAGT 92
Db 1327 LysAspSerGluGlyArgAsnMetValPheProTyrProGlyThrArgAlaLysGluGlu 1346
QY 91 CTCAGT----- 86
Db 1347 LeuSerGlyAlaArgAspSerSerSerTrpGluArgGlnAlaProProThrGlnProLeu 1366
QY 85 -----TCTTTCATCTCCAAAATGGGAACAATCTTTTCG 53
Db 1367 GlyLysGluThrGluSerLeuGlyAlaGlyPheValValMetGlyValAspLeuSer 1386
QY 52 GTCCGGGAAGGGCTTCTGAGGCCAGGGAATCTCTGTTGCTGCCCC 2
Db 1387 -----ArgCysGlyProGluHisProAlaSerArgCysPro 1398
```

Search completed: September 21, 2004, 15:27:21
Job time : 177 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:41:06 ; Search time 461 Seconds
(without alignments)
6213.685 Million cell updates/sec

Title: US-10-776-827-10

Perfect score: 8026

Sequence: 1 cggggcagcaaccaggagat.....aattaaaggtagctgc 4460

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 2684796

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODSL=frame+n2p.model -DEV=xlsp
-Q=/cgn2_1/USPTO_spool_p/US10776827/runat_21092004_133243_20889/app_query.fasta_1.4615
-DB=Published Applications AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10776827@cgn_1.1.724 @runat_21092004_133243_20889
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description

1	507.5	6.3	99	12	US-10-243-552-565	Sequence 565, App
2	170	2.1	88	14	US-10-106-698-6499	Sequence 6499, App
3	159	2.0	28	12	US-10-243-552-951	Sequence 951, App
4	127.5	1.6	1203	14	US-10-174-677-103	Sequence 103, App
5	127.5	1.6	1203	15	US-10-080-334-192	Sequence 192, App
6	121.5	1.5	1251	12	US-10-114-270-80	Sequence 80, App
7	121	1.5	314	16	US-10-437-963-136195	Sequence 136195, App
8	117	1.5	314	16	US-10-767-701-42364	Sequence 42364, App
9	115	1.4	1747	16	US-10-437-963-168997	Sequence 168997, App
10	113.5	1.4	451	9	US-09-965-528-13	Sequence 13, App
11	113.5	1.4	451	10	US-09-946-374-142	Sequence 142, App
12	113.5	1.4	451	12	US-10-206-915-282	Sequence 282, App
13	113.5	1.4	451	12	US-10-139-670-282	Sequence 282, App
14	113.5	1.4	451	12	US-10-201-858-282	Sequence 282, App
15	113.5	1.4	451	12	US-09-969-984-13	Sequence 13, App
16	113.5	1.4	451	12	US-10-205-890-282	Sequence 282, App
17	113.5	1.4	451	12	US-10-208-024-282	Sequence 282, App
18	113.5	1.4	451	12	US-10-201-853-282	Sequence 282, App
19	113.5	1.4	451	12	US-10-063-745-82	Sequence 82, App
20	113.5	1.4	451	12	US-10-063-512-82	Sequence 82, App
21	113.5	1.4	451	12	US-10-063-513-82	Sequence 82, App
22	113.5	1.4	451	12	US-10-063-515-82	Sequence 82, App
23	113.5	1.4	451	12	US-10-063-549-82	Sequence 82, App
24	113.5	1.4	451	12	US-10-063-551-82	Sequence 82, App
25	113.5	1.4	451	12	US-10-174-581-282	Sequence 282, App
26	113.5	1.4	451	12	US-10-176-483-282	Sequence 282, App
27	113.5	1.4	451	12	US-10-176-749-282	Sequence 282, App
28	113.5	1.4	451	12	US-10-176-914-282	Sequence 282, App
29	113.5	1.4	451	12	US-10-176-915-282	Sequence 282, App
30	113.5	1.4	451	12	US-10-066-485A-142	Sequence 142, App
31	113.5	1.4	451	12	US-10-013-907A-142	Sequence 142, App
32	113.5	1.4	451	12	US-10-015-499A-142	Sequence 142, App
33	113.5	1.4	451	12	US-10-063-555-82	Sequence 82, App
34	113.5	1.4	451	12	US-10-063-563-82	Sequence 82, App
35	113.5	1.4	451	12	US-10-063-594-82	Sequence 82, App
36	113.5	1.4	451	12	US-10-063-554-82	Sequence 82, App
37	113.5	1.4	451	12	US-10-180-550-282	Sequence 282, App
38	113.5	1.4	451	12	US-10-183-014-282	Sequence 282, App
39	113.5	1.4	451	12	US-10-187-738-282	Sequence 282, App
40	113.5	1.4	451	12	US-10-187-740-282	Sequence 282, App
41	113.5	1.4	451	12	US-10-187-883-282	Sequence 282, App
42	113.5	1.4	451	12	US-10-194-363-282	Sequence 282, App
43	113.5	1.4	451	12		
44	113.5	1.4	451	12		
45	113.5	1.4	451	12		

ALIGNMENTS

RESULT 1

US-10-243-552-565
; Sequence 565, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: Polypeptides
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623

```

; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 565
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-565

Alignment Scores:
Pred. No.: 1,33e-38 Length: 99
Score: 507.50 Matches: 96
Percent Similarity: 96.97% Conservatives: 0
Best Local Similarity: 96.97% Mismatches: 0
Query Match: 6.32% Indels: 3
DB: 12 Gaps: 1

US-10-776-827-10 (1-4460) x US-10-243-552-565 (1-99)

QY 442 ATGGAGAGATCAGTGCCTCTCAGCTCTATCTGGGACACCATCTTGACCAACACCAA 501
Db 1 MetGluArgIleSerAlaPheSerSerIleTrrPaspThrIleuLeuThrLysHisGln 20

QY 502 GAAGCATCTACAACACCATCTGCCTGGGAGTCCTCTGGGCTGCCACTTTGGTGATC 561
Db 21 GluGlyIleTyrAsnThrIleCysLeuGlyValLeuLeuGlyLeuProLeuValIle 40

QY 562 ATCACACTCTCTCATCTGTGTGCCATTGCTGTGAGCCACCAGGAGGGGCCAG 621
Db 41 IleThrLeuLeuPheIleCysHisCysTrrSerProGlyLysArgGlyGln 60

QY 622 CACCCAGAG-----AGAAAGAGAGAGAGAGAGAGAGATGAAGAGACCTC 672
Db 61 GlnProGluLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80

QY 673 TGGATCTCTCAACCCCAAGCTTCTCCAGATGGAGAGAGAGACCATCATGCTGTT 729
Db 81 TrpIleSerAlaGlnProLysLeuLeuGlnMetGluLysArgProSerLeuProVal 99

RESULT 2
US-10-106-698-6499
; Sequence 6499, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6499
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6499

; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 951
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-951

Alignment Scores:
Pred. No.: 6,71e-06 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 12 Gaps: 0

US-10-776-827-10 (1-4460) x US-10-243-552-951 (1-28)

Alignment Scores:
Pred. No.: 1.01e-06 Length: 88
Score: 170.00 Matches: 36
Percent Similarity: 97.37% Conservatives: 1
Best Local Similarity: 94.74% Mismatches: 1
Query Match: 2.12% Indels: 1
DB: 14 Gaps: 0

US-10-776-827-10 (1-4460) x US-10-106-698-6499 (1-88)

QY 2844 GTGAAGACATCTCCCACTGCTCACCAGCAAGTGCCTTCTCTGCTC 2903
Db 51 MetGluAspIleSerProGlnLeuProLeuLeuThrArgThrSerCysProSerCysLeu 70

QY 2904 CACCTCTCAGTCCCTCCCTAGATGGATGGCTGG-GGAGAGGTGGAGGCTGACAGC 2956
Db 71 HisLeuSerValProLeuGluTrpMetAlaGlyGlyGluValGluAlaAspSer 88

RESULT 3
US-10-243-552-951
; Sequence 951, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 951
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-951
```

1. **Author(s)** _____

QY	3932	GTTACTTCTTAACATTAGGTATTACCCGCACATCAGACATCACTCAAAATGAT---	CAA	3988
Db	1156	-----ProThrSerArgAsnSerLeuGln	1164	
QY	3989	TGCAGGAGCTCTTTCTGTGACACAAATGTCCAGCCCTCCCT-----	4030	
Db	1165	ArgSerGluProTrpCys-----ValProAlaIleProAlaThrGlnGluAlaGlu	1181	
QY	4031	---GGTCACCGCCTTCGCCATGGTAGAGTCGTAGGCTGAGGATGAGGATATGGCTGTC	4087	
Db	1182	AlaGlyGluPheLeuGluProGlyArg-----ArgArgLeuGlnAsnHisSerPro	1198	
QY	4088	TCACCCCTTG	4096	
Db	1199	SerAlaLeu	1201	

```

RESULT 7
US-10-437-963-136195
; Sequence 136195, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 136195
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_37798C.1.pep
; US-10-437-963-136195

```

Alignment Scores:		
Pred. No.:	0.0747	Length: 314
Score:	121.00	Matches: 47
Percent Similarity:	43.4%	Conservative: 29
Best Local Similarity:	26.8%	Mismatches: 57
Query Match:	1.51%	Indels: 42
DB:	16	Gaps: 10
US-10-776-827-10 (1-4460) x US-10-437-963-136195 (1-314)		

US-10-776-827-10 (1-4460) x US-10-437-963-136195 (1-314)

QY	151	TC	TGAATGCACAGGGGGGGTGTTCAGGATTTGCTTACTAGCTGTGACGTGACCTCCAGGA	210
Db	66	Ser	GLuLeuGLyAspAluLeuIleSerProCysMetCysLysGlyThr	81
QY	211	AAG	CAGTCTCTGGCCGAGATCC---CCTGACAAACGCAAGCAAGAAATACGTGGAAGAG	267
Db	82	Gln	GlnPheValHisArgSerCysLeuAspHis	95
QY	268	GT	TCCCAAGCTGGTGGCCATTTCTGCTGTGTGGAGGTGCTGTCAGTGGGATGCC	327
Db	96	Val	LysGlu---GlyPheAlaPheSerHisCysThrThrCysAlaGlnPheHisLeu	114
QY	328	CAA	ACCACCAAGCTGGGAAGGAATAATACAAAGTGGTCAAGTTGCATCTCTTTTGAGCT	387
Db	115	Arg	ValGluThrTrpGluAspAsnSerTrpArgLysMetLysPheArgIlePheValAla	134
QY	388	CAG	GACTGCTTGAAGCCGAGGGTTCCTCGCCCTAATCTAGCAACAGCACCATG---	444
Db	135	Arg	AspValIleLeuVal-----PheLeuAlaValGlnLeuProSerAlaMetIle	151

```

445  -----GAGAGAATCAGTGCCTTC-----TTCAGCTCTATC 474
152 GlyAlaIleSerTyrPheLeuAspArgAspGlySerPheArgAsnSerPheSerAspGly 171
475 TGGGACACCATCTTGACCAAAACACCAAGAGGATCTACAAACCATCTCGCTGGGAGTC 534
172 TrpAspArgPheLeuSerLysHisProIleProPheTyr-----TyrCysIleGlyVal 189
535 CTCCTGGGCTGCACATCTTGGTCATCATCACACTCTCTCTCATCTGT----- 582
190 ValValPhePheValLeuLeuGlyPhePheGlyLeuIleLeuHisCysSerSerPheAsn 209
583  -----TGC-----CATGTGCTGTGG 597
210 AspAsnGlnAspProCysLeuAlaGlyCysArgAsnCysCysTyr 224

RESULT 8
US-10-767-701-42364
; Sequence 42364, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 42364
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C61842_1.pep
US-10-767-701-42364

```

Alignment Scores:	
Pred. NO.:	0.178
Score:	117.00
Percent Similarity:	42.61%
Best Local Similarity:	25.57%
Query Match:	1.46%
ns	16
Length:	314
Matches:	45
Conservative:	30
Mismatches:	57
Indels:	44
Gaps:	10

```

US-10-776-827-10 (1-4460) x US-10-767-701-42364 (1-314)

QY      151 TCTGAATCCACAGGCGGGTGTTCAGCGATTGTTTACTACGTTGAACGTGACTCCTCCAGGA   210
        ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      66 SerGIuIIeGlyAspGIuLeuIleSerProCysMetCysLysGlyThr-----   81

QY      211 AAGCAATGTCGCCGAGATCC---CCTGCACAACGCAAGAAGATACCTGGAAAGGAG   267
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      82 GIuGIuPheValHisArgSerCysLeuAspHis-----TriArgSer   95

QY      268 GTCCTCCCAAGTCGCTGCCATTITGCTGCTGTGTGTGGAGGTGCTGTCACTGGCATGCC   327
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db      96 ValLysGIu---GlyPheAlaPheSerHisCysThrThrCysLysAlaGIuPheHisIeu   114

QY      328 CAACCCCAAAGCTGGAAGAGGAATAAATACAAGTGTCGAAGTTGCATCCTTTTGAGCT   387
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     115 ArgValGIuThrTrpGIuAspAsnSerTrpHisLysMetLysPheArgIlePheValAla   134

QY      388 CAGGACCTGCTT-----GYAAGCGGAGAGGTTCCTCTG   420
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     135 ArgAspValLeuValPheLeuAlaValGIuLeuThrIleAlaIleGlyAlaIle   154

QY      421 GCCTTAATCTAGCCAAGCACCTGGAGAGCAANTCAGTGCCCTTC-----TTCACTCT   471
        :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db     155 AlaTyrrPhe-----LeuaspArgAspGlySerPheArgAsnSerPheSerAsp   170

```

```

QY 472 ATCTGGGACACCATCTTTGACCAACACCAAGAGGCATCTACACACCATCTGCTGGGA 531
Db 171 GlyTyrAspArgPheLeuSerIleHisProIleProPheTyr-----TyrCysIleGly 188
QY 532 GTCTCTCTGGGCTGCGCATCTTTGGTGATCATCACACTCTCTTCATCTGT----- 582
Db 189 ValValValPhePheValLeuLeuGlyPhePheGlyLeuIleValHisCysSerSerPhe 208
QY 583 -----TGC-----CATCTGCTGCTGG 597
Db 209 AsnAspAsnGlnAspProCysLeuAlaGlyCysArgAsnCysCysTyr 224

RESULT 9
US-10-437-963-168997
; Sequence 168997, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 168997
; LENGTH: 1747
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67458C.1.pdp
US-10-437-963-168997

Alignment Scores:
Pred. No.: 0.58 Length: 1747
Score: 115.00 Matches: 77
Percent Similarity: 33.45% Conservative: 22
Best Local Similarity: 26.01% Mismatches: 109
Query Match: 1.43% Indels: 88
DB: 16 Gaps: 14

US-10-776-827-10 (1-4460) x US-10-437-963-168997 (1-1747)
QY 160 ACAGGGCGGCTTTCAGCGATTGTTTACTACGTTCAGCTGACCTCCAGGAAGAGCTTC 219
Db 1056 ThrProGlnValValSerAlaLeuVal-ValGluArgAspGluAspAsnProHisSe 1075
QY 220 TGGCCGA---GATCCCTGCACAAACCAAGCAAGAAAGTAACGTGGAAGAGGCTCCCAA 276
Db 1075 rAlaHisProHisProValSerThrArgProGlyArgGluGlnGlyGluAlaProGI 1095
QY 277 GCTGCTGCGCATTTTGTCTGTGTGTGGAGGTCT----- 313
Db 1095 uProAsnGlyGlyProArgProProThrAlaGlyAlaGlyProProProAlaCysProTh 1115
QY 314 -GTCAGTGCGCATGCCCAACCAACCAAGCTGGAAGAGGATTAATTAACAAGTGTCAAGTT 372
Db 1115 rValProGlyAlaProAspProPolyAspGlyProGlyAlaThrAla-----GlyArgPr 1133
QY 373 GCATCTTTTGTAGCTCAGGACCT-----GCTTGTAAAGCCGAGAG 411
Db 1133 oHis-LeuLeuProSerAspProGluValValGlyThrGluAlaGluCysAlaProArg 1153
QY 412 GGTCTCTGGCCCTTAATCTAGCCAAGCACCATTGGAGAGATCATGTCCTTCTTCAGCTCT 471
Db 1153 lyLeuSerAspGluGlu----- 1158

```

```

QY 472 ATCTGGGACACCATCTTTGACCAACACCAAGAGGCATCTA-----CAACACCATCTGC 525
Db 1159 -----HisProValSerThrArgProGlyArgGluGlnGlyGluAlaProGluP 1176
QY 526 CTGGGAGTCTCTGCGGCTGCGCATCTTTGGTGATCATCACACTCTCTTCATCTCTTGC 585
Db 1176 rAsnGlyGlyProArgProProThrAlaGlyAlaGlyProLeuProAla---CysLeuT 1195
QY 586 CATTGCTGCTGGAGCCCAACAGGAGGCGGCGAGCCAGCAGCAGAGAGAGAGAGAGAGAG 645
Db 1195 hrValProGlyAlaProAspProGlnAspGlyProGlyAlaThrAla----- 1210
QY 646 AAGAAGAAGAGATGAAGAAGACCTCTGGAATCTCTGTCACACCCCAAGCTTCTCCAGATG 705
Db 1211 -----GlyArgPro-----ArgLeuSerProSerA 1219
QY 706 GAGAAGAGACCATCACTGCTGCTTTAGTTAGGACGAGAGCAGAGGTTTCTTCTTGGG 765
Db 1219 spProGlu-----ValValGlyThrGluAlaGluCysAlaPro----- 1231
QY 766 GCTAAGCCTCTTCTGACCAACACAGACATTTTCAGGAACCCCTGAAATATATGCACTATG 825
Db 1232 -----ArgGlyLeuSerAsp----- 1236
QY 826 TCATGTCACAGAGTAATCTACTCAACCAAGAGAAACAACCTCAGACTAAGTGTCCAGTG 885
Db 1237 --GluGluHisProValSerThrArgProGlyArgGluGlnGlyGluAlaProSerA 1256
QY 886 GAGGCGAGTCCCGAGGACCAAGTGGACATTTCTCGATACTCTTGGAGTACTCTTGGCAGTATGTGTC 945
Db 1256 rgThrAlaAlaGlnGlyProArgArgProGluLeu-----AlaLeuCysP 1271
QY 946 CAATAGCAATGCTCTTACTGACAGACCCAGGATGCTCCCTCCACCC 989
Db 1271 rLeuValArgArgCysGlnGlu-----ProProThr 1281

RESULT 10
US-09-965-528-13
; Sequence 13, Application US/09965528
; Publication No. US20020187523A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: LAL, Preeti
; APPLICANT: BURFORD, Neil
; APPLICANT: BANDMAN, Olga
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: PATTERSON, Chandra
; TITLE OF INVENTION: EXTRACELLULAR SIGNALING MOLECULES
; FILE REFERENCE: PF-0701 USA
; CURRENT APPLICATION NUMBER: US/09/965,528
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/134,949
; PRIOR FILING DATE: 1999-05-19
; PRIOR APPLICATION NUMBER: 60/144,270
; PRIOR FILING DATE: 1999-07-15
; PRIOR APPLICATION NUMBER: 60/146,700
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/157,508
; PRIOR FILING DATE: 1999-10-04
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: incyte ID No. US20020187523A1 2267403CD1
US-09-965-528-13

```

Alignment Scores:

Assignment Score:			
Pred. No.:	0.447	Length:	45
Score:	113.50	Matches:	90
Percent Similarity:	33.16%	Conservative:	38
Best Local Similarity:	23.32%	Mismatches:	124
Best Local Similarity:	23.32%	Matches:	134
Query Match:	1.41%	Indels:	13
DR:	9	Gaps:	23

US-10-776-827-10 (1-4460) x US-09-965-528-13 (1-451)

	Qy	318	GTGCGATGCCAACCCAAAGCTGGGAAGGAATAAATTACAACTGGTTCACAGTTGCATC	377
	Db	79	VallHisCysProGlnPro-----ValThrGluProGlnGlnCysCys	92
	Qy	378	CTTTTGAGCTCAGGACTGCTTGT---AAGCCGAGAGGGTTCTCTGGC---	422
	Db	93	-----ProlYsCysValGluProHisThrProSerGlyLeuArgAlaPro	107
	Qy	423	CCTTAATCTATGCCCAAGCAC-----CATGGAGAG-	449
	Db	108	ProlYsSerCysGlnHisAsnGlyThrMet-Tyr-GlnHisGlyGluIlePheSerAlaHis	127
	Qy	450	-----NATCATGGCTTCTT- - - - -	467
	Db	128	GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGlyGln	147
	Qy	468	CTCTATCTGGGACACCATTCTTGACAAACACCAAGAAGGCATCTACAACACCATCTGCCT	527
	Db	148	IlefyrCysGly-----LeuThrThrCysProGlu-----Pro	158
	Qy	528	GGGAGTCTCTCTGGCGCTGCCACTCTTGTGTGATCATCACACTCTCTTCATCTGTGTGCCA	587
	Db	159	GlyCysPro-----AlaProLeuProLeuProAsp	168
	Qy	588	TTCGTCTGGAGCCACCACGACAGAGGGCGGACGACAGAGAAAAGAAAGAA	647
	Db	169	SerCysCysGlnAlaCysLysAspGlu-----AlaSerGluGlnSerAspGluGlu	185
	Qy	648	GAAGAA-----GAAGGATGAAGAGACCTCTGGATCTCTGCTCAACCCAAAGCT	695
	Db	186	AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla	205
	Qy	696	TCTCCAGATGGAGAAGACCATCATCGCTGCTTTTAGTTAGGACGAGGACGAGAGGTGTTT	755
	Db	206	Gly-Arg--LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeuS	224
	Qy	756	CTTTTCTGGGCTAAGCCTCTCTTGACCACACAGACATTTTCAGAAACCCCTGAAATA	815
	Db	224	erPheIle-----ProArgHisPheArg-----	231
	Qy	816	ATGCACTATGTCCATGTCCACAGAGTAATACTACTCAAC-	857
	Db	232	-----ProlYsGlyAlaGlySerThrThrVallysileValLeuLysG	246
	Qy	858	AACAAACCTCAGACTAAGTGTCCCAGTGGAGGAGTCCCCAGGACCCACGTGGACAAATC	917
	Db	246	lu-LysHisLysLysAlaCysValHisGlyLysThrTyrSer---HisGlyGluVal	264
	Qy	918	TTGGATATGTCTTGGCATATGTGTCCAATA--GCAATGCTCTTCTATCTCAGACCCA	974
	Db	265	TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluasp	284
	Qy	975	GGCATGCCCTCCC-----ACCTGTCTCTGGCATACCCACATGCAAGCACAAA	1022
	Db	285	GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro--CysArgHisPro	303
	Qy	1023	GAA-----CAT	1028
	Db	304	GluLysValAlaGlyLysCysCysLysIleCysProGluAspLysAlaAspProGlyHis	323
	Qy	1029	TTATCCATCATCTCTCAATATGGTTCCCAAGTGTGTGCATGTCAGCTAACACACACAC	1088

RESULT 11

```

US-09-346-374-142
Sequence 142, Application US/09946374
Publication No. US20030073129A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Pat, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tunas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmitted
TITLE OF INVENTION: Acids Encoding
FILE REFERENCE: P2830PlC1
CURRENT APPLICATION NUMBER: US/09/99946374
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/098716
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098723
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098749
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098750
PRIOR FILING DATE: 1998-09-01
PRIOR APPLICATION NUMBER: 60/098803
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098821
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/098843
PRIOR FILING DATE: 1998-09-02
PRIOR APPLICATION NUMBER: 60/099536
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099602
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099642
PRIOR FILING DATE: 1998-09-09

```

1	PRIOR APPLICATION NUMBER: 60/099741	2	PRIOR FILING DATE: 1998-09-24
2	PRIOR FILING DATE: 1998-09-10	3	PRIOR APPLICATION NUMBER: 60/101743
3	PRIOR APPLICATION NUMBER: 60/099754	4	PRIOR FILING DATE: 1998-09-24
4	PRIOR FILING DATE: 1998-09-10	5	PRIOR APPLICATION NUMBER: 60/101915
5	PRIOR APPLICATION NUMBER: 60/099763	6	PRIOR FILING DATE: 1998-09-24
6	PRIOR FILING DATE: 1998-09-10	7	PRIOR APPLICATION NUMBER: 60/101916
7	PRIOR APPLICATION NUMBER: 60/099792	8	PRIOR FILING DATE: 1998-09-24
8	PRIOR FILING DATE: 1998-09-10	9	PRIOR APPLICATION NUMBER: 60/102207
9	PRIOR APPLICATION NUMBER: 60/099808	10	PRIOR FILING DATE: 1998-09-29
10	PRIOR FILING DATE: 1998-09-10	11	PRIOR APPLICATION NUMBER: 60/102240
11	PRIOR APPLICATION NUMBER: 60/099812	12	PRIOR FILING DATE: 1998-09-29
12	PRIOR FILING DATE: 1998-09-10	13	PRIOR APPLICATION NUMBER: 60/102307
13	PRIOR APPLICATION NUMBER: 60/099815	14	PRIOR FILING DATE: 1998-09-29
14	PRIOR FILING DATE: 1998-09-10	15	PRIOR APPLICATION NUMBER: 60/102330
15	PRIOR APPLICATION NUMBER: 60/099816	16	PRIOR FILING DATE: 1998-09-29
16	PRIOR FILING DATE: 1998-09-10	17	PRIOR APPLICATION NUMBER: 60/102331
17	PRIOR APPLICATION NUMBER: 60/100385	18	PRIOR FILING DATE: 1998-09-29
18	PRIOR FILING DATE: 1998-09-15	19	PRIOR APPLICATION NUMBER: 60/102484
19	PRIOR APPLICATION NUMBER: 60/100388	20	PRIOR FILING DATE: 1998-09-30
20	PRIOR FILING DATE: 1998-09-15	21	PRIOR APPLICATION NUMBER: 60/102487
21	PRIOR APPLICATION NUMBER: 60/100390	22	PRIOR FILING DATE: 1998-09-30
22	PRIOR FILING DATE: 1998-09-15	23	PRIOR APPLICATION NUMBER: 60/102570
23	PRIOR APPLICATION NUMBER: 60/100584	24	PRIOR FILING DATE: 1998-09-30
24	PRIOR FILING DATE: 1998-09-16	25	PRIOR APPLICATION NUMBER: 60/102571
25	PRIOR APPLICATION NUMBER: 60/100627	26	PRIOR FILING DATE: 1998-09-30
26	PRIOR FILING DATE: 1998-09-16	27	PRIOR APPLICATION NUMBER: 60/102684
27	PRIOR APPLICATION NUMBER: 60/100661	28	PRIOR FILING DATE: 1998-10-01
28	PRIOR FILING DATE: 1998-09-16	29	PRIOR APPLICATION NUMBER: 60/102687
29	PRIOR APPLICATION NUMBER: 60/100662	30	PRIOR FILING DATE: 1998-10-01
30	PRIOR FILING DATE: 1998-09-16	31	PRIOR APPLICATION NUMBER: 60/102965
31	PRIOR APPLICATION NUMBER: 60/100664	32	PRIOR FILING DATE: 1998-10-02
32	PRIOR FILING DATE: 1998-09-16	33	PRIOR APPLICATION NUMBER: 60/103258
33	PRIOR APPLICATION NUMBER: 60/100683	34	PRIOR FILING DATE: 1998-10-06
34	PRIOR FILING DATE: 1998-09-17	35	PRIOR APPLICATION NUMBER: 60/103314
35	PRIOR APPLICATION NUMBER: 60/100684	36	PRIOR FILING DATE: 1998-10-07
36	PRIOR FILING DATE: 1998-09-17	37	PRIOR APPLICATION NUMBER: 60/103315
37	PRIOR APPLICATION NUMBER: 60/100710	38	PRIOR FILING DATE: 1998-10-07
38	PRIOR FILING DATE: 1998-09-17	39	PRIOR APPLICATION NUMBER: 60/103328
39	PRIOR APPLICATION NUMBER: 60/100711	40	PRIOR FILING DATE: 1998-10-07
40	PRIOR FILING DATE: 1998-09-17	41	PRIOR APPLICATION NUMBER: 60/103395
41	PRIOR APPLICATION NUMBER: 60/100848	42	PRIOR FILING DATE: 1998-10-07
42	PRIOR FILING DATE: 1998-09-18	43	PRIOR APPLICATION NUMBER: 60/103449
43	PRIOR APPLICATION NUMBER: 60/100849	44	PRIOR FILING DATE: 1998-10-06
44	PRIOR FILING DATE: 1998-09-18	45	PRIOR APPLICATION NUMBER: 60/103633
45	PRIOR APPLICATION NUMBER: 60/100919	46	PRIOR FILING DATE: 1998-10-08
46	PRIOR FILING DATE: 1998-09-17	47	PRIOR APPLICATION NUMBER: 60/103678
47	PRIOR APPLICATION NUMBER: 60/100930	48	PRIOR FILING DATE: 1998-10-08
48	PRIOR FILING DATE: 1998-09-17	49	PRIOR APPLICATION NUMBER: 60/103679
49	PRIOR APPLICATION NUMBER: 60/101014	50	PRIOR FILING DATE: 1998-10-08
50	PRIOR FILING DATE: 1998-09-18	51	PRIOR APPLICATION NUMBER: 60/103711
51	PRIOR APPLICATION NUMBER: 60/101068	52	PRIOR FILING DATE: 1998-10-08
52	PRIOR FILING DATE: 1998-09-18	53	PRIOR APPLICATION NUMBER: 60/104257
53	PRIOR APPLICATION NUMBER: 60/101071	54	PRIOR FILING DATE: 1998-10-14
54	PRIOR FILING DATE: 1998-09-18	55	PRIOR APPLICATION NUMBER: 60/104987
55	PRIOR APPLICATION NUMBER: 60/101279	56	PRIOR FILING DATE: 1998-10-20
56	PRIOR FILING DATE: 1998-09-22	57	PRIOR APPLICATION NUMBER: 60/105000
57	PRIOR APPLICATION NUMBER: 60/101471	58	PRIOR FILING DATE: 1998-10-20
58	PRIOR FILING DATE: 1998-09-23	59	PRIOR APPLICATION NUMBER: 60/105002
59	PRIOR APPLICATION NUMBER: 60/101472	60	PRIOR FILING DATE: 1998-10-20
60	PRIOR FILING DATE: 1998-09-23	61	PRIOR APPLICATION NUMBER: 60/105104
61	PRIOR APPLICATION NUMBER: 60/101474	62	PRIOR FILING DATE: 1998-10-21
62	PRIOR FILING DATE: 1998-09-23	63	PRIOR APPLICATION NUMBER: 60/105169
63	PRIOR APPLICATION NUMBER: 60/101475	64	PRIOR FILING DATE: 1998-10-22
64	PRIOR FILING DATE: 1998-09-23	65	PRIOR APPLICATION NUMBER: 60/105266
65	PRIOR APPLICATION NUMBER: 60/101476	66	PRIOR FILING DATE: 1998-10-22
66	PRIOR FILING DATE: 1998-09-23	67	PRIOR APPLICATION NUMBER: 60/105693
67	PRIOR APPLICATION NUMBER: 60/101477	68	PRIOR FILING DATE: 1998-10-26
68	PRIOR FILING DATE: 1998-09-23	69	PRIOR APPLICATION NUMBER: 60/105694
69	PRIOR APPLICATION NUMBER: 60/101738		

10

Page 10

Percent Similarity:	23.32%
Best Local Similarity:	23.32%
Mismatches:	124

Alignment Scores:		
Pred. No.:	0.447	Length: 451
Score:	113.50	Matches: 90
Percent Similarity:	33.16%	Conservative: 38
Best Local Similarity:	23.32%	Mismatches: 124

Query Match:	1.41%	Indels:	134
DB:	12	Gaps:	23
US-10-776-827-10 (1-4460) x US-10-206-915-282 (1-451)			
QY	318	GTGGCATGCCCAACCCAAAGCTGGAAGAGGAATAAAATACAAAGTGTCAAGTGTGCATC	377
DB	79	ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys	92
QY	378	CTTTGAGCTCAGGACCTGCTTGT---AAGCGAGAGGGTCTCTGGC-----	422
DB	93	-----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro	107
QY	423	CCTAATCTAGCAACAC-----CATGGAGAG-----	449
DB	108	ProLysSerCysGlnHisAsnGlyThrMetTyrGlnHisGlyGluIlePheSerAlaHis	127
QY	450	-----AATCAGTGCCTCTT-----CAG	467
DB	128	GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGln	147
QY	468	CTCTATCTGGACACCATCTTGCACCAACACCAAGAGGCATCTCAACACCATCTGCT	527
DB	148	IleTyrCysGly-----LeuThrThrCysProGlu-----Pro	158
QY	528	GGGAGTCTCTCGGCGCTGCTCTTGTGTGATCATCACTCTCTCTCTCTCTCTCTCT	587
DB	159	GlyCysPro-----AlaProLeuProLeuProAsp	168
QY	588	TTGCTGTGGAGCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	647
DB	169	SerCysCysGlnAlaCysIleAspGlu-----AlaSerGluGlnSerAspGlu	185
QY	648	GAAGAA-----GAAGGATGAAGAGACCTCTGGATCTCTGCTCAACCAAGCT	695
DB	186	AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAsp	205
QY	696	TCTCAGATGAGAGAGACCATCACTGCTGTTAGTTAGTGGCAGGAGGAGGAGGAGG	755
DB	206	Gly-Arg-----LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAla	224
QY	756	CTTTCTGGGGCTAAGCCTCTCTGACACACACAGACATTTACAGGAACCCCTCAATA	815
DB	224	erPheIle-----ProArgHisPheArg-----	231
QY	816	ATGCACTATGCTCATGTCACAGAGTAATCTCAACC-----AAGG	857
DB	232	-----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG	246
QY	858	AACAAACCTCAGACTAAGTGTCCAGTGGAGGAGTCCAGGAGGAGGAGGAGGAGGAG	917
DB	246	lu-LysHisLysLysAlaCysValHisGlyGlyLysThrTyrSer---HisGlyGlu	264
QY	918	TTGGATACTGTCTTGGCAGCTATGTGTCCAATA---GCAATGCTCTCTTACTGCAC	974
DB	265	TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGlu	284
QY	975	GGCATGCTCCC-----ACCTGTCTCTGGCATACCCCAATGCAAGACACAAA	1022
DB	285	GlyArgGlnAspCysGlnArgValThrCysProThrGluTyrPro---CysArgHis	303
QY	1023	GAA-----CAT	1028
DB	304	GluLysValAlaGlyLysCysCysLysIleCysProGluAspLysAlaAspProGly	323
QY	1029	TTATCCATCATCTCAATATGTTCCCAAGTGTGTGCACATGCAGGTAAACACACAC	1088
DB	324	SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThr	343
QY	1089	ACAAATTCAGTAGCAGTACGTGGGCAAGTATATCTCTCATCAATGCTCATGCT	1148
DB	344	ValSerProSerProAspAsnLeuArgArgPheAlaLeu-----	356

QY	1149	ATGTACTTGTGCAGGAGTAGTATCTACAGTCACAAAATGTCTCATCGGAAAGCC	1208
DB	357	-----GluHisGluAlaSerAspLeuValGluIleTyrLeuTrpLysLeu	371
QY	1209	TTGCAGATTCAGACACATATACAAATTTCTTCAACCAAGGAGGAGGAGGAGGAG	1262
DB	372	ValLysAspGluThrGluAlaGlnArgGlyGluValProGlyProArgProHis	391
QY	1263	COATCTATTCCA	1274
DB	392	GlnAsnLeuPro	395
RESULT 13			
US-10-199-670-282			
; Sequence 282, Application US/10199670			
; Publication No. US20040033560A1			
; GENERAL INFORMATION:			
; APPLICANT: Baker, Kevin P.			
; APPLICANT: Chen, Jian			
; APPLICANT: Desnoyers, Luc			
; APPLICANT: Goddard, Audrey			
; APPLICANT: Godowski, Paul J.			
; APPLICANT: Gurney, Austin L.			
; APPLICANT: Pan, James			
; APPLICANT: Smith, Victoria			
; APPLICANT: Watanabe, Colin K.			
; APPLICANT: Wood, William I.			
; APPLICANT: Zhang, Zemin			
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC			
; FILE REFERENCE: P3430R1C401			
; CURRENT APPLICATION NUMBER: US/10/199,670			
; PRIOR FILING DATE: 2002-07-19			
; PRIOR APPLICATION NUMBER: 10/052586			
; PRIOR FILING DATE: 2002-01-15			
; PRIOR APPLICATION NUMBER: 60/059263			
; PRIOR FILING DATE: 1997-09-18			
; PRIOR APPLICATION NUMBER: 60/059266			
; PRIOR FILING DATE: 1997-09-18			
; PRIOR APPLICATION NUMBER: 60/062250			
; PRIOR FILING DATE: 1997-10-17			
; PRIOR APPLICATION NUMBER: 60/063120			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: 60/063121			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: 60/063486			
; PRIOR FILING DATE: 1997-10-21			
; PRIOR APPLICATION NUMBER: 60/063540			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR APPLICATION NUMBER: 60/063541			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR APPLICATION NUMBER: 60/063544			
; PRIOR FILING DATE: 1997-10-28			
; PRIOR Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 612			
; SEQ ID NO 282			
; LENGTH: 451			
; TYPE: PRT			
; ORGANISM: Homo Sapien			
US-10-199-670-282			
Alignment Scores:			
Pred. No.: 0.447			
Score: 113.50			
Length: 451			
Matches: 90			
Percent Similarity: 33.16%			
Best Local Similarity: 23.32%			
Query Match: 1.41%			
DB: 12			
US-10-776-827-10 (1-4460) x US-10-199-670-282 (1-451)			
QY	318	GTGGCATGCCCAACCCAAAGCTGGAAGAGGAATAAAATACAAAGTGTCAAGTGTGCATC	377

Db	206	Gly-Arg---	LySArgLyProGlyThrProAlaProThrGlyLeu-SerAlaProLeuS	224
Qy	756	CGTTTCTGGGGCTAAGCCCTCTCTGACCACACAGACATTTTCAGGAACCCCTGAATA	815	
Db	224	erPheile-----	ProArgHisPheArg-----	231
Qy	816	ATGCACATATGTCATGTCTCCACAGAGTAACACTCTCAACC-----	AAGG	857
Db	232	-----	ProLySGlyAlaGlySerThrThrVallysillevallLeuLySG	246
Qy	858	AACAACCTCAGACTAAGTGTCTCCAGTCGAGGGCAGTCCCGAGGCCACGCTGACGAATTC	917	
Db	246	lu-LySHisLySylsAlaCySValHisGlyGlyLysThrTySer---HisGlyGluVal	264	
Qy	918	TTGGATACTGCTTTGGCAGCTATGTGTCCAATA--GCAATGTCTCTTACTCTCAGACCCA	974	
Db	265	TrpHisProAlaPheArgAlaPheGlyProLeuProCysAlaIleuCysThrCysGluAsp	284	
Qy	975	GGCATGCGCTCCC-----	ACTGTCTCTGGCATACCCCATCCCAATGCAAGGCACAA	1022
Db	285	GlyArgGlnAspCysGlnArgValThrCysProThrGluTyPro---CysArgHisPro	303	
Qy	1023	GAA-----	CAT	1028
Db	304	GluLySValAlaGlyLysCysCysLysIleCysProGluAspLySAlaAspProGlyHis	323	
Qy	1029	TTATCCATACATCTCAATATATGTTTCCCAAGTGTGTGCATCATGCATGCAACACACACACAC	1088	
Db	324	SerGluIleSerSerThrArgCysProLySAlaProGlyArgValLeuValHisThrSer	343	
Qy	1089	ACAAATTCAGGTACAGGTACGTGGGCAAGTATATCTGCTCATCAAAATGGTCAATGGCT	1148	
Db	344	ValSerProSerProAspAsnLeuArgPheAlaLeu-----	356	
Qy	1149	ATGTACTTTGTGCAGGNAAGTACATTATCTACAGTCACAAAATGTCTCATGGAAAGCC	1208	
Db	357	-----	GluHisGluAlaSerAspLeuValGluIleTyLeuTrpLyLeu	371
Qy	1209	TTGCCAGATTCCAGACACATATATCAAAATTTCTTAACGCAAGC-----	CCCATACA	1262
Db	372	VallysAspGluThrGluAlaGlnArgGlyGluValProGlyProArgProHisSer	391	
Qy	1263	CCATCTATCCA	1274	
Db	392	GlnAsnLeuPro	395	

Search completed: September 21, 2004, 16:53:47
Job time : 498 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:10:50 ; Search time 72.5 Seconds
(without alignments)
6351.775 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaacaggagat.....aattaaaggtagctgc 4460

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xl
-Q=/cgn2_1/USPTO_spool_p/US10776827/runat_21092004_133241_20811/app_query.fasta_1.4615
-DB=Issued Patents AA -QFMT=fastran -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCI=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn_1_119@runat_21092004_133241_20811 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pcp:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pcp:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pcp:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pcp:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pcp:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	113.5	1.4	631	3	US-08-927-219-127
2	113	1.4	630	3	US-08-927-219-2
3	113	1.4	630	3	US-08-927-219-4
4	109	1.4	180	4	US-09-252-991A-31030
5	109	1.4	408	4	US-09-798-051-6
6	109	1.4	429	4	US-09-798-051-5
7	108.5	1.4	763	4	US-09-252-991A-30146
8	107	1.3	281	4	US-09-252-991A-23962
9	106.5	1.3	2842	1	US-07-741-940-7
10	106.5	1.3	2842	1	US-08-289-548A-7
11	106.5	1.3	2842	1	US-08-452-654-7
12	106.5	1.3	2842	4	US-08-449-731-7

13	106.5	1.3	2843	1	US-07-741-940-2	Sequence 2, Appli
14	106.5	1.3	2843	1	US-08-289-548A-2	Sequence 2, Appli
15	106.5	1.3	2843	1	US-08-452-654-2	Sequence 2, Appli
16	106.5	1.3	2843	1	US-08-452-655B-2	Sequence 2, Appli
17	106.5	1.3	2843	1	US-08-452-655B-7	Sequence 7, Appli
18	106.5	1.3	2843	2	US-08-370-235A-2	Sequence 2, Appli
19	106.5	1.3	2843	3	US-08-450-582-2	Sequence 2, Appli
20	106.5	1.3	2843	3	US-08-450-582-7	Sequence 2, Appli
21	106.5	1.3	2843	4	US-08-449-731-2	Sequence 7, Appli
22	106.5	1.3	2973	2	US-08-821-355A-7	Sequence 7, Appli
23	106.5	1.3	2973	2	US-09-003-687A-7	Sequence 7, Appli
24	106.5	1.3	2973	3	US-09-136-605-7	Sequence 7, Appli
25	105	1.3	1086	4	US-09-252-991A-25051	Sequence 25051, A
26	104.5	1.3	243	4	US-09-252-991A-32776	Sequence 32776, A
27	104	1.3	503	4	US-09-252-991A-18052	Sequence 18052, A
28	104	1.3	1184	4	US-09-266-225D-18	Sequence 18, Appli
29	104	1.3	5405	3	US-08-718-388-9	Sequence 9, Appli
30	104	1.3	530	4	US-09-800-729-112	Sequence 112, App
31	101.5	1.3	1185	3	US-09-041-886-23	Sequence 23, Appli
32	101	1.3	569	4	US-09-252-991A-24513	Sequence 24513, A
33	101	1.3	1940	2	US-08-644-271-30	Sequence 30, Appli
34	101	1.3	1940	1	US-09-077-955-34	Sequence 34, Appli
35	100.5	1.3	1187	1	US-08-320-559-28	Sequence 28, Appli
36	100.5	1.3	1187	3	US-08-545-860D-28	Sequence 28, Appli
37	100.5	1.3	1187	5	PCT-US94-04496-28	Sequence 28, Appli
38	100.5	1.3	1210	1	US-08-320-559-26	Sequence 26, Appli
39	100.5	1.3	1210	3	US-08-545-860D-26	Sequence 26, Appli
40	100.5	1.3	1210	5	PCT-US94-04496-26	Sequence 26, Appli
41	100	1.2	1257	1	US-08-340-428B-49	Sequence 49, Appli
42	99	1.2	224	4	US-09-252-991A-23849	Sequence 23849, A
43	98.5	1.2	558	4	US-09-252-991A-23318	Sequence 23318, A
44	98	1.2	431	4	US-09-252-991A-20267	Sequence 20267, A
45	98	1.2	589	1	US-07-668-648-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-927-219-127
; Sequence 127, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:

APPLICANT: Bell, Graeme I.
APPLICANT: Yamagata, Kazuya
APPLICANT: Oda, Naohisa
APPLICANT: Kaisaki, Pamela J.
APPLICANT: Furuta, Hiroto
APPLICANT: Horikawa, Yukio
APPLICANT: Menzel, Stephen
TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,219
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,679
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:

Wed Sep 22 17:04:10 2004

```

; APPLICATION NUMBER: US 60/028,056
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-927-219-127

Alignment Scores:
Pred. No.: 0.0169 Length: 631
Score: 113.50 Matches: 103
Percent Similarity: 31.76% Conservative: 45
Best Local Similarity: 22.10% Mismatches: 179
Query Match: 1.41% Indels: 139
DB: 23 Gaps: 23

US-10-776-827-10 (1-4460) x US-08-927-219-127 (1-631)
QY 225 GAGATCCCTGTACACGCAAGAGAGTACGTGGAGGAGGCTCCCAAGCTGGCTG 284
D 193 GluLeuPro---ThrIysLysGlyArgArgAsnArgPheLysTrpGlyProAlaSerGln 211
QY 285 GCCATTTCGTGCTGTGTGAGGTGCTGTAGTGGCATGCCAAACCCAAAGCTGGAA 344
D 212 GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu 227
QY 345 GAGGAATAAATTACAAAGTGTGAGGTGCTCATCTCTTTTGTAGCTCAGGACCTGTTGTAAG 404
D 228 GluArg--GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
QY 405 CCAGAGGGTTCTCTGGCCCTTAATCTAGCAAGCACCACCATGGAGAGATCAGTGCCTTCTT 464
D 243 GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGAGGATCTACAAACCATCTG 524
D 262 ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu 281
QY 525 -----CCTGGAGTCTCTCTGGGCTGCTGGGCTGCCACTCTTGGT 557
D 282 AlaMetAspThrTyrSerGlyProProGlyProGlyProGlyProAla---LeuPro 300
QY 558 GATCATCAGCTCTCTCTCTGTTGTCCTGCTGAGCCGACCCAGGAGAGAGGG 617
D 301 AlaHisSerSerPro---GlyLeuProProAlaLeuSerProSerLysValHisGly 319
QY 618 -----CCAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGATGA 662
D 320 ValArgTyrGlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly--- 338
QY 663 AGAAGACCTCTGGATCTCTGCTCAACC-----CAAGCTTCCAGATGAGAGAGAG 713
D 339 ---GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlu 357
QY 714 ACCATCAGCTGCTGTTAGTAGCAGGAGCAGAGGTGTTCTCTTCTGGGGTAAGCC 773
D 358 Pro---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 374
QY 774 TCCTTCTGACCACACAGACATTCAGGAACCCCTGAAATGATGCACTATGTCATGTC 833
D 375 -----GlyProLeuProValSerThrLeuThrAlaLeu 386

```

```

834 CACAGA-----GTAACACTCAACCAAGAAACAACCTCAGACT 872
D 387 HisSerLeuGluGlnThrSerProGlyLeuAsnGlnProGlnAsnLeuMetAla 406
QY 873 AAGTGTCCTCCAGTGGAGGCGAGTCCAGGAGCA----- 905
D 407 SerLeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThr 425
QY 905 ----- 905
D 426 PheThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSer 445
QY 906 -----CGTGGACAATTCCTGGATATCTGCTTGGCAGCT----- 938
D 446 ValProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSer 465
QY 939 -----ATGTGTCCAATAGCAATGCTCTCTTACT 965
D 466 GlnProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThr 485
QY 966 GCAGACCCAGGATGCTCCACCTGTCTCTGGCATACCCACATGCA-----AAGCAC 1019
D 486 GlnAsnProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHis 505
QY 1020 AAGAAACATTTATCC-----ATACATCTCAATATGTTCCCAAGTGTGTGCACATGAC 1073
D 506 LysProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeu 523
QY 1074 GTAACACACACACACAAATTCAGTAGCAGGTAGGCGAAGTATATCTGCTCATC 1133
D 524 IleThrAspThr----- 527
QY 1134 AATGGTCTATTGGCTATGTACTTTGTGCGAGGAAGTACATATCTACAGTCACAAAATG 1193
D 528 -----ThrAsnLeuSerAlaLeuAlaSerLeu 536
QY 1194 TCTCATGGAAAGCCTTGCAGATTCAGACACATATATACAAATTCCTAACGACGAGGC 1253
D 537 ThrProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGly 554
QY 1254 CCCCATACACATCTATTCCATAAACCACTCAGGTACAGATGCATGCTTCTCTTCTTCT 1313
D 555 LeuHisThrProAlaSerGlnAlaThrThrLeuHisValProSerGlnAspProAlaGly 574
QY 1314 -----AACTCTACACATAAATCTTACTGGAAGTACTCATAATTTGGACATTC 1361
D 575 IleGlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSer 594
QY 1362 AGCAACCTGCTACAGTCC 1379
D 595 LeuValLeuTyrGlnSer 600

RESULT 2
US-08-927-219-2
; Sequence 2, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Katsuki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```

STATE: Texas
 COUNTRY: USA
 ZIP: 77210
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/927,219
 FILING DATE: Concurrently Herewith
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/029,679
 FILING DATE: 30-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/028,056
 FILING DATE: 02-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/025,719
 FILING DATE: 10-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, Mark B.
 REGISTRATION NUMBER: 37,259
 REFERENCE/DOCKET NUMBER: ARCD:272
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7577
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 630 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-927-219-2

Alignment Scores:
 Pred. No.: 0.019 Length: 630
 Score: 113.00 Matches: 103
 Percent Similarity: 31.6% Conservative: 44
 Best Local Similarity: 22.15% Mismatches: 180
 Query Match: 1.41% Indels: 138
 DB: 3 Gaps: 23

US-10-776-827-10 (1-4460) x US-08-927-219-2 (1-630)

QY 225 GAGATCCCTGACACGCAAGCAAGAGTAACTGGAGGAGGCTCCCAAGCTGGCTG 284
 D 193 GluLeuPro---ThrLysLysGlyArgArgAsnArgPheLysTyrGlyProAlaSerGln 211
 QY 285 GCCATTTCGTGCTGTGTGGAGTGCTGTCACTGGGATGCCCAACCCAAAGCTGGAA 344
 D 212 GlnLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu 227
 QY 345 GAGGAATAAATAACAAGTGGTCAAGTTGTCATCTTTGAGTCTCAGGACCTGCTGTGAAG 404
 D 228 GluArg---GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
 QY 405 CCGAGAGGTTCTCTGGCCCTTAATCTAGCCAAAGCACCATGAGAGAAATCACTGCTTCTT 464
 D 243 GlnArgGlyValSer---ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
 QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGCTCTACCAACACCATCTG 524
 D 262 ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu 281
 QY 525 -----CCTGGAGTCTCTCTGGGCTGCTGGGCTGCTGGCTGCTGGT 557
 D 282 AlaMetAspThrTyrSerGlyProProProGlyProGlyProGlyProAla---LeuPro 300
 QY 558 GATCATCACATCTCTTTCATCTGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 617
 D 301 AlaHisSerSerPro---GlyLeuProProProAlaLeuSerProSerLysValHisGly 319

QY 618 -----COAGCAGCCAG 665
 D 320 ValArgGlyGlnProAlaThrSerGluThrAlaGluValProSerSerGly----- 337
 QY 666 AGACCTCTCGATCTCTGCTCAACC-----CAAGCTTCTCCAGATGGAGAGAGACC 716
 D 338 GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGluPro 357
 QY 717 ATCACTGCTCTGTTTAGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 776
 D 358 ---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 373
 QY 777 TTCTGACCACACAGACATTTTCAGGAACCCCTGAATAATGCATCTATCTCCATGTGCCAC 836
 D 374 -----GlyProLeuProProValSerThrLeuThrAlaLeuHis 386
 QY 837 AGA-----GTAACACTCTCAACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 875
 D 387 SerLeuGluGlnThrSerProGlyLeuAsnGlnProGlnAsnLeuLeuMetAlaSer 406
 QY 876 TGTCCAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
 D 407 LeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThrPhe 425
 QY 905 ----- 905
 D 426 ThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSerVal 445
 QY 906 -----CGTGGACAATCTTGGATCTGCTGGCAGCT----- 938
 D 446 ProValIleAsnSerMetGlySerSerLeuThrThrLeuGlnProValGlnPheSerGln 465
 QY 939 -----ATGTGTCCTCAATAGCAATGCTCTTACTGCA 968
 D 466 ProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThrGln 485
 QY 969 GACCCAGGACATGCTCCACCTGCTCTGGCATACCCACCATGCA-----AAGCACAAA 1022
 D 486 SerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHisLys 505
 QY 1023 GAACATTTATCC-----ATACATCTCAATATGTTCCCAAGTGTGTCACATGACGTA 1076
 D 506 ProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeuIle 523
 QY 1077 ACACACACACACACAAATTCAGGTAGCAGGTACGTGGGCAAGTATATTCTGCTCATCAA 1136
 D 524 ThrAspThr----- 526
 QY 1137 TGGTCATTGGCTATGTACTTTTGTGAGGAGAGTACATATCTACAGTCACAAAATGTCT 1196
 D 527 -----ThrAsnLeuSerAlaLeuAlaSerLeuThr 536
 QY 1197 CATGGGAAGCTTCCAGATTCAGACACATATATACAAATTTCTACACCAAGGCCCC 1256
 D 537 ProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGlyLeu 554
 QY 1257 CATACACATCTATTTCCATAAACCACTCAGGTTACAGATGCATGCTTCTCTATTCTTCT 1313
 D 555 HisThrProAlaSerGlnAlaThrLeuHisValProSerGlnAspProAlaGlyIle 574
 QY 1314 -----AACTACACATAAACTTTTACTGGAGTACTCATATAATTGGACATTTCCAGC 1364
 D 575 GlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSerLeu 594
 QY 1365 AACCTGCTACAGTCC 1379
 D 595 ValLeuTyrGlnSer 599

RESULT 3
 US-08-927-219-4
 ; Sequence 4, Application US/08927219
 ; Patent No. 6187533

QY 1257 CATACACCATCTATTCATTAACCACTCAGGTTACAGATGATGCTTTCTCTATTTCT--- 1313
 Db |||||
 555 HisThrProAlaSerGlnAlaThrThrLeuHisValProSerGlnAspProAlaGlyIle 574
 QY 1314 -----AACTCTACACATAAATCTTTTACTGGAAGTACTATAATTGGACATTCCAGC 1364
 Db |||||
 575 GlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSerLeu 594
 QY 1365 AACCTGCTACAGTCC 1379
 Db |||||
 595 ValLeuTyrGlnSer 599

RESULT 4
 US-09-252-991A-31030
 ; Sequence 31030, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 31030
 ; LENGTH: 180
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-31030

Alignment Scores:
 Pred. No.: 0.0233 Length: 180
 Score: 109.00 Matches: 50
 Percent Similarity: 39.46% Conservative: 23
 Best Local Similarity: 27.03% Mismatches: 62
 Query Match: 1.36% Indels: 50
 DB: Gaps: 4

US-10-776-827-10 (1-4460) x US-09-252-991A-31030 (1-180)
 QY 206 CAGAAAGCAGTCTGGCCGAGATCCCTGCAACCGCAAGCAAGTAAGTGAAGG 265
 Db |||||
 29 GlnGlySerArgAlaGlySerGlnProLeuSerLysGlnAlaSerSer----- 44
 QY 266 AGGCTCCCAAGCTGGCTGGCCATTTTGTCTGTGTGTGGAGGTGCTGTCAGTGGCATG 325
 Db |||||
 45 -----ArgSerAlaSerHisThrAspSerAlaSerThrGlySerValGly----- 59
 QY 326 CCCAAACCAAGCTGGAAGGAGGAATAATTACAAG---TGCTCAAGTTGTCATCTTTT 382
 Db |||||
 60 -----SerCysGlyArgSerValSerLysGlnAlaArgSerArgValSerSer--- 75
 QY 383 GAGCTCAGGACCTGCTTGAAGCCGAGAGGGTTCTGCGCCCTAATCTAGCCACACCA 442
 Db |||||
 76 -----SerSerArgCysMetSerProIleSerArgSerArgArg 88
 QY 443 TGGAGAGAATCAGTGCCTTCTTCACTCTATCTGGGACACCATCTTGACCAACACCAAG 502
 Db |||||
 89 TrpArgAsnCysSerGlyGlySerSerLeuSerTrpAla-----MetArgArg 104
 QY 503 AAGGATCTACACACCATCTGCTGGGAG----- 532
 Db |||||
 105 ArgAlaSerGlyValArgSerSerTrpAspArgValArgSerSerArgCysTrpAla 124
 QY 533 -----TCCCTCGGCGCTGCCACTCTTGGTGATCATCATCTCCCT 574
 Db |||||
 125 ArgLeuArgArgArgSerAlaMetCysSerArgAlaTrpAlaSerSerProSerSer 144
 QY 575 TCATCTGTGTCATTGCTGTGAGGCCACCAGGCAAGAGGGGCCAGCCAGCAGGAGA 634

Db |||||
 145 SerArgAla---AlaSerGlyAla-----AlaArgGlyValLeuSerSerSer 161
 QY 635 AAAAAGAAAGAAAGAAAGAGGATGAAGAAGACCTCTCGATCTCTGCTCAACCCAAAGC 694
 Db |||||
 162 AlaArgAlaSerAlaArgSerArgLysArgPheSerGlyThr----- 176
 QY 695 TTCTCCAGATGGAGA 709
 Db |||||
 177 ---ThrArgTrpArg 180

RESULT 5
 US-09-798-051-6
 ; Sequence 6, Application US/09798051
 ; Patent No. 6632780
 ; GENERAL INFORMATION:
 ; APPLICANT: Zhang, Ke
 ; APPLICANT: Cam, Linh
 ; APPLICANT: Nakayama, Naoki
 ; TITLE OF INVENTION: Chordin-Like-2 Molecules and Uses Thereof
 ; FILE REFERENCE: 01-005
 ; CURRENT APPLICATION NUMBER: US/09/798,051
 ; CURRENT FILING DATE: 2001-03-05
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 408
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-798-051-6

Alignment Scores:
 Pred. No.: 0.0383 Length: 408
 Score: 109.00 Matches: 90
 Percent Similarity: 31.70% Conservative: 33
 Best Local Similarity: 23.20% Mismatches: 130
 Query Match: 1.36% Indels: 135
 DB: Gaps: 23

US-10-776-827-10 (1-4460) x US-09-798-051-6 (1-408)
 QY 318 GTGGCATGCCCAACCCAAAGCTGGAAGAGGAATAATACAGGTGGTCAAGTTGCATC 377
 Db |||||
 58 ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys 71
 QY 378 CTTTGAGCTCAGGACCTGCTGTGT---AAGCCGAGAGGGTTCTCTGGC----- 422
 Db |||||
 72 -----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro 86
 QY 423 CCTAATCTAGCCAAAGCAC-----CATGGAGAG----- 449
 Db |||||
 87 ProLysSerCysGlnHisAsnGlyThrMetTyrGlnHisGlyGluIlePheSerAlaHis 106
 QY 450 -----AATCAGTGCCTCTTT-----CAG 467
 Db |||||
 107 GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCysThrGluGlyGln 126
 QY 468 CTCTATCTGGACCACTCTTGACCAACACCAAGAGGACATCTACACACCATCTGCCT 527
 Db |||||
 127 IleTyrCysGly-----LeuThrCysProGlu-----Pro 137
 QY 528 GGGAGTCTCTCTGGGCGCTGCCACTCTTGGTGATCATCATCTCTTCTATCTGTGCA 587
 Db |||||
 138 GlyCysPro-----AlaProLeuProLeuProAsp 147
 QY 588 TTGCTGTGGAGCCCAACGAGAGGGCGCCAGCCAGAGAGAAAGAAAGAGAA 647
 Db |||||
 148 SerCysCysGlnAlaCysLysAspGlu-----AlaSerGluGlnSerAspGluGlu 164
 QY 648 GAAGAA-----GAAGGATGAAGAGACCTCTGATCTCTCTCTCAACCCAGCT 695
 Db |||||
 165 AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla 184

Wed Sep 22 17:04:10 2004

```
QY 696 TCTCCAGATGGAGAGACACATCACTGCTGTTTGTAGTGGAGGAGGAGGAGGTTT 755
Db 185 Gly-Arg--LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeuS 203
QY 756 CTTTCTGGGCTAAGCTCTCTGACCACACAGACATTTCCAGGAACCCCTGAAATA 815
Db 203 erPheille-----ProArgHisPheArg----- 210
QY 816 ATGCACATATGTCATGTCACAGAGTACTACTCAACC-----AAGG 857
Db 211 -----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG 225
QY 858 AACAAACCTCAGACTAAGTGTCCAGTGGAGGCGAGTCCCGAGGACCACTGGCAATTC 917
Db 225 lu-LysHisLysLysAlaCysValHisGlyLysThrThrSer---HisGlyGluVal 243
QY 918 TTGGATACTGCTTGGCAGCTATGTGCAATA---GCAATGCTCTTACTGCACACCA 974
Db 244 TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluasp 263
QY 975 GGCATGCTCTCC-----ACCTGTCTCTGGCATACCCCATGCAAGACACAAA 1022
Db 264 GlyArgGlnaspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro 282
QY 1023 GAA-----CAT 1028
Db 283 GluLysValAlaGlyLysCysCysLysIleCysProGluaspLysAlaaspProGlyHis 302
QY 1029 TTATCCATACATCTCAATATGTTCCCAAGTGTGTGCACATGCATTAACACACACAC 1088
Db 303 SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer 322
QY 1089 ACAAAATTCAGGTACGAGTACGTGGCGCAAGTATATCTG----- 1127
Db 323 ValSerProSerProAspAsnLeuArgPheAlaLeuGluHisGluAlaSerAspLeu 342
QY 1128 -----CTCATCAATAGTCAATGGCTATGTTGTGCGAGGAGTACATATCTACA 1181
Db 343 ValGluIleTyrLeuTrpLysLeuValLysGlyIlePheHisLeuThrGlnIleLysLys 362
QY 1182 GTCACAAAATGTCTCATGGGAAGCCTTCCAGATTCCAGACACATATACAAATTTCT 1241
Db 363 ValArgLysGlnaspPheGlnLysGluAlaGln-----HisPheArg 376
QY 1242 AACGAGCAAGGCCCAT 1259
Db 377 LeuLeuAlaGlyProHis 382
```

```
RESULT 6
US-09-798-051-5
; Sequence 5, Application US/09798051
; Patent No. 6632780
; GENERAL INFORMATION:
; APPLICANT: Zhang, Ke
; APPLICANT: Cam, Linh
; APPLICANT: Nakayama, Naoki
; TITLE OF INVENTION: Chordin-Like-2 Molecules and Uses Thereof
; FILE REFERENCE: 01-005
; CURRENT APPLICATION NUMBER: US/09/798,051
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-798-051-5
```

```
Alignment Scores:
Pred. No.: 0.0395 Length: 429
Score: 109.00 Matches: 90
Percent Similarity: 31.70% Conservative: 33
Best Local Similarity: 23.20% Mismatches: 130
```

```
Query Match: 1.36% Indels: 135
DB: Gaps: 23
US-10-776-827-10 (1-4460) x US-09-798-051-5 (1-429)
QY 318 GTGCATGCCCAACCAACCAAGCTGGAAGAGGATAAATTAACAAGTGGTCAAGTTGCATC 377
Db 79 ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys 92
QY 378 CTTTGTAGCTCAGGACCTGCTTGT---AAGCCGAGAGGGTTCTCTGGC----- 422
Db 93 -----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro 107
QY 423 CCTAATCTAGCCAAAGCAC-----CATGGAGAG----- 449
Db 108 ProLysSerCysGlnHisasnGlyThrMetTyrGlnHisGlyGluIlePheSerAlaHis 127
QY 450 -----AATCAGTGCCTTCTT-----CAG 467
Db 128 GluLeuPheProSerArgLeuProasnGlnCysValLeuCysSerCysThrGluGlyGln 147
QY 468 CTCATCTGGGACACCATCTTGACCAACACACAGAGGCATCTACACACCATCTGCT 527
Db 148 IleTyrCysGly-----LeuThrThrCysProGlu-----Pro 158
QY 528 GGGAGTCTCTCTGGGCTGCCACTCTTGGTGATCATCATCTCTCTTCTCATCTGTTGCCA 587
Db 159 GlyCysPro-----AlaProLeuProLeuProAsp 168
QY 588 TTGCTGTGGAGCCACACAGGCAAGAGGGCCAGCAGAGAGAGAGAGAGAGAGAA 647
Db 169 SerCysCysGlnAlaCysLysaspGlu-----AlaSerGluGlnSerSerAspAla 185
QY 648 GAAGAA-----GAAGGATGAAGAAGACCTCTGGATCTCTGTCAACCCCAAGCT 695
Db 186 AspSerValGlnSerLeuHisGlyValArgHisProGlnAspProCysSerSerAspAla 205
QY 696 TCTCCAGATGGAGAGAGACCATCACTGCTGTTTGTAGTGGAGGAGGAGGAGGAGGTTT 755
Db 206 Gly-Arg--LysArgGlyProGlyThrProAlaProThrGlyLeu-SerAlaProLeuS 224
QY 756 CTTTCTGGGCTAAGCTCTCTGACCACACACACATTTCCAGGAACCCCTGAAATA 815
Db 224 erPheille-----ProArgHisPheArg----- 231
QY 816 ATGCACATATGTCATGTCACAGAGTAACTACTCAACCC-----AAGG 857
Db 232 -----ProLysGlyAlaGlySerThrThrValLysIleValLeuLysG 246
QY 858 AACAAACCTCAGACTAAGTGTCCAGTGGAGGCGAGTCCCGAGGACCACTGGCAATTC 917
Db 246 lu-LysHisLysLysAlaCysValHisGlyLysThrThrSer---HisGlyGluVal 264
QY 918 TTGGATACTGCTTGGCAGCTATGTGTCGAATA---GCAATGCTCTTACTGCAGACCCA 974
Db 265 TrpHisProAlaPheArgAlaPheGlyProLeuProCysIleLeuCysThrCysGluasp 284
QY 975 GGCATGCTCTCC-----ACCTGTCTCTGGCATACCCCATGCAAGACACAAA 1022
Db 285 GlyArgGlnaspCysGlnArgValThrCysProThrGluTyrPro---CysArgHisPro 303
QY 1023 GAA-----CAT 1028
Db 304 GluLysValAlaGlyLysCysCysLysIleCysProGluaspLysAlaaspProGlyHis 323
QY 1029 TTATCCATACATCTCAATATGTTCCCAAGTGTGTGCACATGCATTAACACACACAC 1088
Db 324 SerGluIleSerSerThrArgCysProLysAlaProGlyArgValLeuValHisThrSer 343
QY 1089 ACAAAATTCAGGTACGAGTACGTGGCGCAAGTATATCTG----- 1127
Db 344 ValSerProSerProAspAsnLeuArgPheAlaLeuGluHisGluAlaSerAspLeu 363
```


QY 3401 ----TCCATGTTTATGCTTCTGATTCCTCTGCTGCACACAGCC---CC 3453
Db 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
QY 3454 AGCCCTTTATCTCTCCATTTTCATCTCTCCAGCTCTGCTCCCTGAACTGCCTGGC 3513
Db 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
QY 3514 AACCCATGGGACC 3526
Db 2493 rThrHisSerSer 2497

RESULT 10
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-289-548A-7

Alignment Scores:
Pred. No.: 0.228
Score: 106.50
Percent Similarity: 35.41%
Best Local Similarity: 24.59%
Query Match: 1.33%
DB: 14
Length: 2842
Matches: 75
Conservative: 33
Mismatch: 105
Indels: 92
Gaps: 14

US-10-776-827-10 (1-4460) x US-08-289-548A-7 (1-2842)
QY 2797 TCACATAACAGAAATAACTTGGCCATCTGCTGCACCAACCCAGGGATGTGGAAGACATCT 2856
Db 2224 SerArgGlyArgThrMetIleHisIleProGlyValArgAsnSerSerSerThrSer 2243
QY 2857 CCCCAACTGCTGCTCACCAGAGCAAGCTGCCCTTCTGCTCCACCTCTCTAGTCC 2916
Db 2244 ProValSerLysGlyProPro-----LeuLysThrProAlaSerLysSer 2259
QY 2917 CCTAGATGGATGGCTGGGAGAGGTGGAGCTGACAGCTGACAGCTAGTGTACATAT 2976
Db 2260 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2279
QY 2977 GATCTA-----GGAGGGGGGATC----- 2994
Db 2280 GluLeuSerProValAlaArgGlnThrSerGlnIleGlyGlySerSerLysAlaProSer 2299
QY 2995 ---ACGGGATCCGGGACCATACAAAGTACATGGTTTCCATGGCAACTGCTTCTGCTTT 3051
Db 2300 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2312
QY 3052 GAATTAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3099
Db 2313 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2329
QY 3100 CCAGGCACAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3147
Db 2330 ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2349
QY 3148 CCCAGGGATGAACACCCAGAACTCCTCTCT----- 3180
Db 2350 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2369
QY 3181 ---CAGTCACTCCACAGCCGATGACTCAGAGAGCCAAACCCAGAAAGGGGCTCTCTCT 3237
Db 2370 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2389
QY 3238 TTCCCATCACAGACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3297
Db 2390 IlePro-----ArgSerGluSerAla 2396
QY 3298 GGATAGGCCCTTAACGTTTGT-----TTAAATAACAGGTGCAT----- 3336
Db 2397 SerLysGlyLeuAsnGlnMetAsnGlyAlaAsnLysLysValGluLeuSer 2416
QY 3337 -----GAAAGGAGCCTAAGGCCATTGTTG 3360
Db 2417 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2436
QY 3361 ATA-----TCCACT 3400
Db 2437 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuGl 2456
QY 3401 ----TCCATGTTTATGCTTCTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3453
Db 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
QY 3454 AGCCCTTTATCT 3513
Db 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
QY 3514 AACCCATGGGACC 3526
Db 2493 rThrHisSerSer 2497

RESULT 11
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS

2330 ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2349
 3148 CCCAGGATGAACACACCCAGAACTCACTTCT 3180
 2350 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2369
 3181 --CAGTCACCTCCACAGCCGATGACTCAGAGAGCCAAACCCAGATGGGGCTCTCTT 3237
 2370 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2389
 3238 TTCCCATCAGAGACTCCCTCGACACCTTTCTCTGGCGTAACCTAGAGAGCTCCACAGTGCA 3297
 2390 IlePro-----ArgSerGluSerAla 2396
 3298 GGATAGGCCCTAAACGTTTGG-----TTAATAAACAGGTGCAT----- 3336
 2397 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2416
 3337 -----GAAAGGAGCCTAAGGCCATTGTTG 2436
 2417 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2436
 3361 ATA-----TCCACTCTCTTCTTCCACTTCCCTCTCTCATCTTTTTC----- 3400
 2437 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuGl 2456
 3401 ----TCCATGTTTATGCTCTCTCTGATTCCCTCTTCTGCTGCACACAGACCGCC---CC 3453
 2456 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2473
 3454 AGCCCTTATTCCTCTCCATTTCCTCTCTCCCTTCCAGCTCTGTCCTGAACCTGCCACTGCC 3513
 2473 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2493
 3514 AACCCATGGGACC 3526
 2493 rThrHisSerSer 2497

RESULT 13
 US-07-741-940-2
 ; Sequence 2, Application US/07741940
 ; Patent No. 5352775
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, GEOFF
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERTIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 94
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner, Birch, McKie & Beckett
 ; STREET: 1001 G Street, NW
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20001-4598
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/741,940
 ; FILING DATE: 19920109
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

NAME:	Kagan, Sarah A.
REGISTRATION NUMBER:	32,141
REFERENCE/DOCKET NUMBER:	1107.035574
TELECOMMUNICATION INFORMATION:	
TELEPHONE:	202-508-9100
TELEFAX:	202-508-9299
INFORMATION FOR SEQ ID NO:	2:
SEQUENCE CHARACTERISTICS:	
LENGTH:	2843 amino acids
TYPE:	AMINO ACID
TOPOLOGY:	linear
MOLECULE TYPE:	protein
US-07-741-940-2	
Alignment Scores:	
Pred. No.:	0.228
Score:	106.50
Percent Similarity:	35.41%
Best Local Similarity:	24.59%
Query Match:	1.33%
DB:	1
US-10-776-827-10 (1-4460) x US-07-741-940-2 (1-2843)	
QY	2797 TCACATAACAGAAATAACTTGCATCTGCTGCACCAAAACCCAGGATGTGGAAGACATCT 2856
Db	2225 SerArgGlyArgThrMetIleHisIleProGlyValAlaArgAsnSerSerSerThrSer 2244
QY	2857 CCCCAACACTGCCATGCTCCACAGAGCAAGCTGCCCTTCTCTCCACCTCTCACTCC 2916
Db	2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY	2917 CCTAGAAATGGATGGCTGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTGCAGATAT 2976
Db	2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY	2977 GATCTA-----GGAGGGCGGATC----- 2994
Db	2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlyLysSerSerLysAlaProSer 2300
QY	2995 ---ACGGGATCCGGACACCTACAGTAACATGTTTCATGGCAACTGCTTGCCTGTTT 3051
Db	2301 ArgSerGlySerArgAspSerThrProSerA-gProAla----- 2313
QY	3052 GAATTAAGACAGACGTGATGTTCATGTCATGACAAAGCCCT-----CTACT 3099
Db	2314 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2330
QY	3100 CCAGGCACAATGCTCCGTGCTCTCTAATCCATGAGCTGCTC-----TCA 3147
Db	2331 ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
QY	3148 CCCAGGATGAACACCCAGAACTCACTTCT----- 3180
Db	2351 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerThrThrSerProGly 2370
QY	3181 ---CAGTCACCTCCACAGCCGATGACATCAGAGAGCAACCCAGATGGGCTCTCTT 3237
Db	2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
QY	3238 TTCCCATCAGACTCCCTGACACCTTCTCTGGCGTAACCTAGAGGAGTCCAGTGCA 3297
Db	2391 IlePro-----ArgSerGluSerAla 2397
QY	3298 GGATAGGCCCTAAAGCTTTTG-----TTAATAAACAGGTGCAT----- 3336
Db	2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAlaAsnLysLysValGluLeuSer 2417
QY	3337 -----GAAAGGACCTTAAGCCATGTTG 3360
Db	2418 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2437

```

QY 2797 TCACATAACAGATAACTTCCCATCTGCTCCGACCAACCCAGGATGTGAAGACATCT 2856
Db 2225 SerArgGlyArgThrMetIleHisProGlyValArgAsnSerSerSerThrSer 2244
QY 2857 CCCCACAACTGCCACTGCTCACCAGGACAAAGTCCCTCTCTCCACCTCTCAGTCC 2916
Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY 2917 CCTAGAAATGGATGGCTGGGAGAGGTGGAGGTGACAGCTGACAGCTGACAGTGTGAT 2976
Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY 2977 GATCTA-----GGAGGGCGGATC----- 2994
Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlySerSerLysAlaProSer 2300
QY 2995 ---ACCGGATCCGGACCACTAACATGTTTCCATGGCAACTGCTGCTGCTT 3051
Db 2301 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2313
QY 3052 GAATTAAGACAGCAGTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 3099
Db 2314 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2330
QY 3100 CCAGGCACAAATGCTCCTGCTGCTCTCTCAATCCATGACTGCTGCTC-----TCA 3147
Db 2331 ProGlyArgAsnGlyLeuSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
QY 3148 CCCCAGGGATGAACACCCAGAACTCACTTCT----- 3180
Db 2351 ProSerThrAlaSerThrLysSerSerGlySerGlyLysMetSerTyrThrSerProGly 2370
QY 3181 ---CAGTCACTTCCACAGCGATGACTAGAGAGCCAAACCCAGAAATGGGCGCTCTCT 3237
Db 2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
QY 3238 TTCCCATCACAGACTCCCTCGACAACTTCTCGCGTAACCTAGAGGAGTCCAGTGA 3297
Db 2391 IlePro-----ArgSerGluSerAla 2397
QY 3298 GGATAGCGCTAAACGTTTGT-----TTAAATAAACAGGTGCAT----- 3336
Db 2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer 2417
QY 3337 -----GAAAGGAGCCTAAGGCCATGTTG 3360
Db 2418 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2437
QY 3361 ATA-----TCCACTCTCTTCTTCCACTTCTCTCTCTCTCTCTCTCTCTCTCT 3400
Db 2438 ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuG 2457
QY 3401 ---TCCAGTTTATGCTTCTCTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3453
Db 2457 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2474
QY 3454 AGCCCTTTATTCCTCTCACTTTCTACTCTCTCCAGCCTCTGTCCTGAACTGCCACTGGC 3513
Db 2474 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2494
QY 3514 AACCCATGGGACC 3526
Db 2494 rThrHisSerSer 2498

```

RESULT 15

```

US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY

```

```

; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-654-2

```

```

Alignment Scores:
Pred. No.: 0.228 Length: 2843
Score: 106.50 Matches: 75
Percent Similarity: 35.41% Conservative: 33
Best Local Similarity: 24.59% Mismatches: 105
Query Match: 1.33% Indels: 92
DB: 1 Gaps: 14

```

```

US-10-776-827-10 (1-4460) x US-08-452-654-2 (1-2843)
QY 2797 TCACATAACAGATAACTTGCCTGCTGCACCAACCCAGGATGTGAAGACATCT 2856
Db 2225 SerArgGlyArgThrMetIleHisProGlyValArgAsnSerSerSerThrSer 2244
QY 2857 CCCCACAACTGCCACTGCTCACCAGGACAAAGTCCCTCTCTCTCCACCTCTCAGTCC 2916
Db 2245 ProValSerLysLysGlyProPro-----LeuLysThrProAlaSerLysSer 2260
QY 2917 CCCTAGAATGGATGGCTGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTGATAT 2976
Db 2261 ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer 2280
QY 2977 GATCTA-----GGAGGGCGGATC----- 2994
Db 2281 GluLeuSerProValAlaArgGlnThrSerGlnIleGlySerSerLysAlaProSer 2300
QY 2995 ---ACCGGATCCGGACCACTAACATGTTTCCATGGCAACTGCTGCTGCTT 3051
Db 2301 ArgSerGlySerArgAspSerThrProSerArgProAla----- 2313

```

```
QY 3052 GAATTAAGACAGACAGTCTGATGTCATGTCCTGATGACCAAGGCTCTCTCTCT 3099
Db 2314 -----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer 2330
QY 3100 CCAGGCACAATGTCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3147
Db 2331 ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer 2350
QY 3148 CCCAGGATGATGAACACCCAGAACTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Db 2351 ProSerThrAlaSerThrLysSerSerSerGlySerGlyLysMetSerTyrThrSerProGly 2370
QY 3181 ---CAGTCACATCCACAGCCGATGACTCAGAGAGCCAAACCCAGAAATGGGCTCTCTCT 3237
Db 2371 ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer 2390
QY 3238 TTCCCCATCAGACTCCCTGACAACTTCTCTGCGCTAACTAGAGGAGTCCCAAGTGCA 3297
Db 2391 IlePro-----ArgSerGluSerAla 2397
QY 3298 GGATAGGCCCTAAACGTTTGTG-----TTAAATAAACAGGTGCAT----- 3336
Db 2398 SerLysGlyLeuAsnGlnMetAsnAsnGlyAlaAsnLysLysValGluLeuSer 2417
QY 3337 -----GAAAGAGCCTAAGGCCATTGTG 3360
Db 2418 ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu 2437
QY 3361 ATA-----TCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3400
Db 2438 ValArgGlnSerThrPheLeuLysGluAlaProSerProThr-LeuArgArgLysLeuGl 2457
QY 3401 ----TCCATGTTTTATGCTTCTCTGATTCCTCTCTCTCTCTCTCTCTCTCTCTCT 3453
Db 2457 uGluSerAlaSerPheGluSerLeuSerProSerSer-----ArgProAlaSerPr 2474
QY 3454 AGCCCTTTATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3513
Db 2474 oThrArgSerGlnAlaGlnThrProValLeuSerProSerLeuProAspMetSerLeuSe 2494
QY 3514 AACCATGGGACC 3526
Db 2494 rThrHisSerSer 2498
```

Search completed: September 21, 2004, 15:48:20
Job time : 196.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:06:59 ; Search time 410 Seconds

(without alignments)
6864.445 Million cell updates/sec

Title: US-10-776-827-10

Perfect score: 8026

Sequence: 1 cggggcagcaacaggagat.....aattaaaggtatactagctc 4460

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xl
-O=/cgn2_1/USPTO_spoop/US10776827/runat_21092004_133240_20778/app_query.fasta_1.4615
-DB=SPTRMBL_25 -QMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plp -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1 1 728 @runat_21092004_133240_20778 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	403	5.0	126	11	Q8CE67	Q8CE67 mus musculus

	2	214.5	2.7	160	11	Q8CBL2
	3	174	2.2	62	4	Q8MW72
	4	171.5	2.1	122	11	Q9D4T5
	5	129.5	1.6	1203	4	Q7Z3U0
	6	127.5	1.6	1203	4	Q9HC56
	7	125.5	1.6	1084	4	Q8N3K7
	8	118	1.5	468	11	Q8BIR7
	9	118	1.5	4998	11	Q8CG65
	10	114.5	1.4	468	11	Q99LH4
	11	114.5	1.4	1381	11	Q8CJ95
	12	114.5	1.4	1400	11	Q8CGB6
	13	114.5	1.4	1637	6	Q9XSV8
	14	114.5	1.4	5146	6	Q8SPM4
	15	113.5	1.4	451	4	Q7Z5J3
	16	112.5	1.4	473	11	Q8VDC2
	17	111.5	1.4	1860	4	Q8IZC6
	18	110	1.4	4988	11	Q8CG65
	19	109	1.4	968	10	Q9SD31
	20	108	1.3	614	13	Q80Z20
	21	108	1.3	1396	5	Q8T0T0
	22	107.5	1.3	786	10	Q9FMJ2
	23	107.5	1.3	1210	11	Q811B5
	24	107	1.3	841	11	Q9WVL8
	25	107	1.3	971	4	Q96KW2
	26	107	1.3	981	11	Q8CHP6
	27	106.5	1.3	964	4	Q8NFT1
	28	106.5	1.3	983	4	Q8NDX5
	29	106.5	1.3	983	4	Q8NFT7
	30	106.5	1.3	1002	11	Q8BP69
	31	106.5	1.3	1055	11	Q8CHG2
	32	106	1.3	382	16	Q8XVU9
	33	106	1.3	655	11	Q8R122
	34	106	1.3	981	11	Q8CBC0
	35	106	1.3	981	11	Q8C881
	36	106	1.3	989	11	Q95ML9
	37	105.5	1.3	1158	5	Q97169
	38	105	1.3	735	4	Q95784
	39	104.5	1.3	1222	11	Q810H4
	40	104	1.3	513	11	Q8R1U3
	41	104	1.3	692	11	Q8BH09
	42	104	1.3	735	11	Q8K4N4
	43	104	1.3	1419	4	Q9NFG0
	44	103.5	1.3	330	4	Q14882
	45	103.5	1.3	1243	12	Q8JKK6

ALIGNMENTS

RESULT 1

Q8CE67
ID Q8CE67 PRELIMINARY; PRT; 126 AA.
AC Q8CE67;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Unclassifiable.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL: AK028922; BAC26196.1; -;
SQ SEQUENCE 126 AA; 14102 MW; CEB01EF3E6C50A44 CRC64;

Alignment Scores:

2 LysThrSerGlySerLeuLeuSerProSerSerSerArgTrpArgArgAspHisHisCys 21

725 CTGTTTAGTTAGCAGAGACAGAGAGTGTTCCTTTCTGGGGCTAAAGCCCTCTCTGACC 784

22 LeuSerSer-GluGlnGlyGlySerValSerSerProTrpSerArgAlaSerPheGlnPr 41

785 ACACACAGACATTTCCAGAACCCCTGAAATATAATGCACATATGTCCATGTCCACAGAGTAAC 844

41 oGlnArg- ---LeuGlyGlyAlaPheAsnValMetHisSerLeu- ---ProGlnI1 57

845 TACTCAACCAAGGAACAAACCTCAGACTAGTGTCCAGTGGAGGCGAGTCCCGAGGACC 904

57 eThrGlnProArgAspLysPheGlnThrThrHisProAsnGly- ---TrpGlyAr 74

905 ACGTGGACAATCTTGATACTCTCTGGCAGCTATGTGTCCAATAGCAATGTCTCTTAC 964

74 gaIrgValAspAsnSerTyrSerValLeuAlaAlaMetCysProValSerVal- --- 91

965 TGCAGACCCAGGCATGCTCCACCTGTCTTGGCATACCCCAATGCAAAGACAAAGA 1024

92 - - - - - pro-GlnAlaTrpThrAlaValSerGlyIleLeuGlnHisLysSerThrGluS 109

1025 ACATTTATCCATAATCATATAGTTGCCAAGTGTGCACATGCACGTAAACACACAC 1084

109 eMetTyrIleCysIle-LeuAspGlyPheGlnLeuCysLeuTyr-HisMet- --- 125

1085 ACACAAATTCAGGTACGACGTAGCTGGCAGATATATCTGTCTCATCAATGTCTATT 1144

126 -TyrAlaHisLeuAspAsnArgTyrMetGlyGlnAspIleLeuPheIleThrGlnProCys 145

1145 GGCTATGTACTTT 1157

145 sAsnIleTyrPhe 149

RESULT 3

Q8W72 PRELIMINARY; PRT; 62 AA.

ID Q8W72

AC Q8W72; 01-MAR-2002 (TRENBLrel. 20, Created)

DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)

DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)

DE Hypothetical protein (Fragment).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC020789; AAH20789.1; -.

KW Hypothetical protein.

FT NON TER 1

SQ SEQUENCE 62 AA; 6816 MW; B57842ABDPF44A82 CRC64;

Alignment Scores:

Pred. No.: 6.72e-09 Length: 62

Score: 174.00 Matches: 35

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 2.17% Indels: 0

DB: 4 Gaps: 0

US-10-776-827-10 (1-4460) x Q8W72 (1-62)

QY 3 GGGCAGCAACACAGAGATTCCTGGGCTTGAGAGCCCTTCGCGCAGCAAGATTG 62

Db 28 GlyGlnGlnProGlyAspSerLeuGlyLeuGlnGluAlaLeuProArgThrGluArgLeu 47

QY 63 TTCCCATTTTGGAGATGAAGAACTGAGCTCAAGCAGCTGAG 107

Db 48 PheProIleLeuGluMetLysLysLeuArgLeuLysAlaAlaGlu 62

Pred. No.:	5,56e-33	Length:	126
Score:	403.00	Matches:	81
Percent Similarity:	78.74%	Conservative:	19
Best Local Similarity:	63.78%	Mismatches:	22
Query Match:	5.02%	Indels:	6
DB:	11	Gaps:	2

US-10-776-827-10 (1-4460) x Q8CE67 (1-126)

QY	442	ATGGAGAGAAATCAGTCCTTTCAGCTATCTGGGACACCATCTTGACCAACACCAA	501
DB	1	MetGluArgAsnSerSerPheAsnSerValTrpGluThrIleAargIleAargHisGlu	20
QY	502	GAGGGATCTACAACACCATCTGCCCTGGGAGTCTCTGGGCTGCGACTCTTGTTGTC	561
DB	21	ArgGlyValPheAsnThrValCysLeuAlaValLeuLeuGlyLeuProLeuValVal	40
QY	562	ATCACATCTCTTCATCTCTTCCATTTGCTGTGGAGCCCAACGAGAGAGGGGCCAG	621
DB	41	LeuThrLeuPhePheIleCysCysHisCysCysSerArgProProLysSerAsnGln	60
QY	622	CAGCCAGAG-----AAGAAAAGAAGAAGAAGAAGATGAAGAAGACCTC	672
DB	61	GlnProAspGlnAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	80
QY	673	TGGATCTCTGCTCAACC-CAAGCTTCTCCAGATGGAGAGAGACCATCATCTGCTGTTTA	731
DB	81	TrpIleSerAlaGlnAlaGlnAlaProAspGlyGluGluThrIleThrAlaCysLeu	100
QY	732	GTTTAGGAGGAGAGAGAGTGTTCCTTCTGGGGCTAAGCTTCTTGACACACACA	791
DB	101	ValSerLysGlyAlaAlaCysProProLeu-GlyAlaGluProProSerAsnHis---	119
QY	792	GACATTTGAGAACCCCTG	810
DB	119	GasPLeuGluGluProLeu	125

RESULT 2

Q8CBL2	PRELIMINARY;	PRT;	160 AA.
ID	Q8CBL2		
AC	Q8CBL2;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Unclassifiable.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;		
RX	MEDLINE=22354683; PubMed=12466851;		
RA	The FANTOM Consortium,		
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;		
RT	"Analysis of the mouse transcriptome based on functional annotation of		
RT	60,770 full-length cDNAs."		
RL	Nature 420:563-573 (2002).		
DR	EMBL; AK035814; BAC29196.1; --		
SQ	SEQUENCE 160 AA; 18109 MW; 9E43A1AB200AAE9C CRC64;		

Alignment Scores:		Length:	160
Pred. No.:	4,91e-13	Matches:	69
Score:	214.50	Conservative:	23
Percent Similarity:	55.09%	Mismatches:	56
Best Local Similarity:	41.32%	Indels:	19
Query Match:	2.67%	Gaps:	5
DB:	11		

US-10-776-827-10 (1-4460) x Q8CBL2 (1-160)

QY	665	RAGACCTTGATCTCTGCTCAACCCAGATCTTCCAGATGGAGAGACCATCATCTGC	724
		

```

RESULT 4
Q9D4T5 PRELIMINARY; PRT; 122 AA.
ID Q9D4T5
AC Q9D4T5
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 4930562F07R1K protein.
GN 4930562F07R1K
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Asburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald J., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016191; BAB30142.1; -.
DR MGD; MG1:1922505; 4930562F07R1K.
SQ SEQUENCE 122 AA; 13568 MW; 514CFF083312AE08 CRC64;

Alignment Scores:
Pred. No.: 1,566-08 Length: 122
Score: 171.50 Matches: 38
Percent Similarity: 69.49% Conservative: 3
Best Local Similarity: 64.41% Mismatches: 17
Query Match: 2.14% Indels: 1
DB: 11 Gaps: 1

US-10-776-827-10 (1-4460) x Q9D4T5 (1-122)
QY 608 CTCTGGTGGG---CTCCAGCAGATGGCAACAGATGAAGAGGAGTGTGATGATCACCAG 552
Db ProGlyGlyArgLeuGlnGlnGlnTrpGlnGlnLysLysLysArgValArgThrThr 82
QY 551 AGTGGAGGCGCCAGGAGGACTCCAGGAGGAGTGTGTGTGTGTGTGTGTGTGTGT 492
Db ArgGlyAsnProAspSerThrAlaArgGlnThrValLeuLysThrProLeuSerCysLeu 102
QY 491 GTCAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 435
Db IleArgMetValSerGlnThrGluLeuLysLysGluLeuPheLeuSerMetValLeu 121

RESULT 5
Q7Z3U0 PRELIMINARY; PRT; 1203 AA.
ID Q7Z3U0
AC Q7Z3U0
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 4930562F07R1K protein.
GN 4930562F07R1K
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Asburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald J., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016191; BAB30142.1; -.
DR MGD; MG1:1922505; 4930562F07R1K.
SQ SEQUENCE 122 AA; 13568 MW; 514CFF083312AE08 CRC64;

Alignment Scores:
Pred. No.: 1,566-08 Length: 122
Score: 171.50 Matches: 38
Percent Similarity: 69.49% Conservative: 3
Best Local Similarity: 64.41% Mismatches: 17
Query Match: 2.14% Indels: 1
DB: 11 Gaps: 1

US-10-776-827-10 (1-4460) x Q9D4T5 (1-122)
QY 608 CTCTGGTGGG---CTCCAGCAGATGGCAACAGATGAAGAGGAGTGTGATGATCACCAG 552
Db ProGlyGlyArgLeuGlnGlnGlnTrpGlnGlnLysLysLysArgValArgThrThr 82
QY 551 AGTGGAGGCGCCAGGAGGACTCCAGGAGGAGTGTGTGTGTGTGTGTGTGTGTGT 492
Db ArgGlyAsnProAspSerThrAlaArgGlnThrValLeuLysThrProLeuSerCysLeu 102
QY 491 GTCAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 435
Db IleArgMetValSerGlnThrGluLeuLysLysGluLeuPheLeuSerMetValLeu 121

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human amygdala;
RA Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA Newes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX537422; CAD97664.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1203 AA; 132214 MW; 25B905E7E6203F10 CRC64;

Alignment Scores:
Pred. No.: 0,000931 Length: 1203
Score: 129.50 Matches: 82
Percent Similarity: 38.23% Conservative: 56
Best Local Similarity: 22.71% Mismatches: 126
Query Match: 1.61% Indels: 97
DB: 4 Gaps: 18

US-10-776-827-10 (1-4460) x Q7Z3U0 (1-1203)
QY 496 CACCAAGAGGAGCATCTACCAACCATCTGCTGGAGTCTCTCTGGGCTGCCACTCTTG 555
Db TyTGlnAsnGluAspTyrLeuThrIleMetIleAlaIleAlaGlyAlaMetValVal 826
QY 556 GTGATC-----ATCACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 594
Db IieValValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla 846
QY 595 -----TGAGCCCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 627
Db AlaGlnArgSerLysGlnGlyAlaGluTyrMetSerProAsnGlnGluAsnLysGlnAsn 866
QY 628 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 681
Db LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 886
QY 682 GCTCAACCCAGGCTCTCCAGATGA-----GAAGAGACCATCACTG 723
Db IleGlu--GluSerLysProAspAlaValHisGluProIleAsnGlyThrIleSerL 906
QY 724 CTCTGTTAGTTAGCAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
Db euAlaAlaGluLeuGluGlnSerIleGlyArgPheAspIleGlySerAlaPro---- 924
QY 781 GACCACACACAGACATTTTCAGGAACCCCTGAATAATGACATATGTCATGTCACACAG 840
Db --ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLysSerA 942
QY 841 TAACCTACTCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db --ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLysSerA 942
QY 942 laSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerValys 959
QY 901 GACCAC-----GTGAGCAATTTCTGGATCTGTCATGTCCTTGGCAGCTA 939
Db LysHisHisValIleGlnGluLeuProLeuAspAsnThrPheValGlyGly- 976
QY 940 TGTGTCCATAGCAATGCTCTTACTGACAGCCAGGAGGAGGAGGAGGAGGAGGAGGAG 990
Db --CysAspThrLeuSer-LysArgSerSerLysArgSerSerSerSerSerSerSerSer 991
QY 991 GTCTCTGGCATACCCAACATGCAAGAGCAAAAGAACATTT-----ATCCA 1035
Db --CysAspThrLeuSer-LysArgSerSerLysArgSerSerSerSerSerSerSerSer 991
QY 1036 TACATCTCAATATGTTTCCCAAGTGTGTGACATGACGCTAACACACACACACACAAATT 1095
Db sThrArgGlnSerGlnArgValThrPheHisLeuProAsp----- 1023
QY 1096 CAGGTAGCAGGTAGTGGGCAAGTATATTCTGCTCATCAATGGTCTATTGGCTATGTA 1155

```


1024	Db	---				:::	GlySerGlnGluSerCysSerAspSerGlyLeuGlyAspHisG1	1038
1156	Qy	TTGTGCAGGGAAGTACATTATCTACAGTCACAAAAATGCT		:::		:::	---	1196
1038	Db	uProValGlySerGlyThrLeuIleSerHis-ProLeuProLeuValGlnProGlnAspG		:::		:::	1058	
1197	Qy	-----CATGGAAAGCCTTCGAGATTCACACACATATATA-----				CAATTTTCCTA	1242	
1058	Db	luPheTyrAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAspProA	:::	:::			1078	
1243	Qy	ACCAGCAAGGCCCCCATACACCACTCTATTCCATAAACCACTCAGGTTACAGATGCATGCT				---	1302	
1078	Db	snSerAspGlyProLeuGlyProArg-----				---	1086	
1303	Qy	TTCTCTATTCTAACTTACACATATAAACTTTTACTGGAAGTACTCATAAATTGGCAATTCCA	:::	:::			1362	
1087	Db	--GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysIleuValLeuGlyHisSerA	:::	:::			1106	
1363	Qy	GCAACTCTCTACAGTCCCCACCCCTTGTTGTGTGATACAGACACACCAAG	:::	:::			1413	
1106	Db	snSerCysTromMetPro---ProGlyLeuGlyProTyrGlnHisProLysI	:::	:::			1121	

RESULT 6		
Q9HC56	PRELIMINARY:	PRT: 1203 AA.
ID	Q9HC56	
AC	Q9HC56	
DT	01-MAR-2001	Created
DT	01-WAR-2001	(TREMBLrel. 16, Created)
DT	01-NOV-2001	(TREMBLrel. 16, Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)
DE	Protocadherin-9.	

OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Kools P.F.J., Van Roy F.;
RT	"Identification and characterization of a novel human protocadherin
RT	gene with high homology to a chromosome X-linked protocadherin.";
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF169692; AAF89689.2; -.
DR	HSSP; P15116; INCU.
DR	Genew; HGNC:8661; PCDH9.
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0005509; F:calcium ion binding; IEA.
DR	GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR	GO; GO:0007155; P:cell adhesion; IEA.
DR	GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR	InterPro; IPR002126; Cadherin.
DR	Pfam; PF00028; cadherin; 6.
DR	PRINTS; PR00205; CADHERIN.
DR	SMART; SM00112; CA; 7.
DR	PROSITE; PS00232; CADHERIN_1; 6.
DR	PROSITE; PS50268; CADHERIN_2; 7.
DR	Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW	SEQUENCE 1203 AA; 132250 MW; CEEDD24FC1CPBSE2 CRK64;
SQ	

```

Alignment Scores:
Pred. No. : 0.00151
Score: 127.50
Percent Similarity: 38.12%
Best Local Similarity: 22.93%
Query Match: 1.59%
DR: 4
Length: 1203
Matches: 83
Conservative: 55
Mismatch: 125
Indels: 99
Gaps: 18

```

[illegible]

QY	556	GTGATC-----ATCACACTCTCTCTTCATCTCTGTTCCTATCTGCTGTC-----	594
Db	827	IleValValIlePheValThrValLeuValArgCysArgHisAlaSerArgPheLysAla	846
QY	595	-----TGGAGCCCCACGAGCAAGAGGCGGCAGCAGCCCA	627
Db	847	AlaGlnArgSerLysGlnGlyAlaGluTrpMetSerProAsnGlnGlnAsnLysGlnAsn	866
QY	628	GAGAGAGAAAAGACAGAGAGAGAGAGGATGAAGACCTC-----TGGATCTCT	681
Db	867	LysLysLysLysArgLysArgLysSerProLysSerSerLeuLeuAsnPheValThr	886
QY	682	GCTCAACCAACAGCTCTCCAGAT-----GGAGAGAGACCA	717
Db	887	IleGlu-----GluSerLysProAspAlaValHisGluProIleAsnGlyThrIleSerL	906
QY	718	TCACCTGCCTGTTTAGTAGGAGAGAGAGAGGTGTTCTCTTCTGGGCTAAGCCCTCT	777
Db	906	euProAlaGluLeuGluGluGlnSerIleGlyArgPheAsp-----TrpGlyProAlaPro-	924
QY	778	TCAGACACACACAGACATTTCCAGGAACCCCTGAAATATGCACTATGTCCTCATGCCA	837
Db	925	-----ProThrThrPheLysProAsnSerProAspLeuAlaLys-----HisTyrLys	941
QY	838	GAGTAACTACTCAACCAAGGAAACAACTCAGACTAAGTCCAGTGGAGGGCAGTCCC	897
Db	941	erAlaSerProGlnPro-AlaPheHisLeuLysProAspThrProVal-----SerVal	958
QY	898	AGGGACCCAC-----GTGGACAATCTTGGATACTGCTTTGGCAG	936
Db	959	LysLysHisHisValIleGlnGluLeuProLeuAsnThrPheValGlyGly-----	976
QY	937	CTATGTGTCCAATAGCAATCTCTTACTCGACACCCAGGCATCCCTCC-----CA	987
Db	977	-----CysAspThrLeuSer-LysArgSerSerThrSerSerAspHis	990
QY	988	CCTGTCTCTGGCATACCCCATCGCAAGACACAAAGAACATTT-----AT	1032
Db	990	sPheSerAlaSer-----GluCysSerSerGlnGlyGlyPheLysThrLysGlyProLe	1008
QY	1033	CCATACATCTCATATATGGTTCCCAAGTGTGTGCACATCCAGTAAACACACACACAA	1092
Db	1008	uHisThrArgGlnSerGlnArgArgValThrPheHisLeuProAsp-----	1023
QY	1093	ATTCCAGTAGCAGTGTGGGCAAGTATATCTGCTCATCAATGGCTATGGCTATGT	1152
Db	1024	-----GlySerGlnLysSerCysSerAspSerGlyLeuGlyAspHis	1037
QY	1153	ACTTTGTGCAGGGAAGTACATTTATCTACAGTCACAAATAATGTCT-----	1196
Db	1037	sGluProValGlySerGlyThrLeuIleSerHis-ProLeuProLeuValGlnProGlnA	1057
QY	1197	-----CATGGAAAGCCTTGCAGATTTCAGACACATATATA-----CAATTTC	1239
Db	1057	spGluPheTyrAspGlnAlaSerProAspLysArgThrGluAlaAspGlyAsnSerAsp	1077
QY	1240	CTAACCCAGCAGGCCCCATACACCATCTATTCCATAAACCACTCAGGTTACAGATGCAT	1299
Db	1077	roAsnSerAspGlyProLeuGlyProArg-----	1086
QY	1300	GCTTTCTCTATTCTAACTCTACACATAAACTTTTCTGGAAGTACTCATAAATTGGACATT	1359
Db	1087	-----GlyLeuAlaGluAlaThrGluMetCysThrGlnGluCysLeuValLeuGlyHis	1105
QY	1360	CCAGCAACTGTGTAGTCCCCCAGCCCTGTGTCTTGTATACAGACACACCAAG	1413
Db	1105	erAspAsn-CysTrpMetPro--ProGlyLeuGlyProTyrGlnHisProLys	1121
RESULT 7			
ID	Q8N3K7	PRELIMINARY;	PRT; 1084 AA.
DT	01-OCT-2002 (TrEMBLRel. 22, Created)		

Db	806	LysIysHisHisValIleGlnGluProLeuAspAsnThrPheValGlyGly----	823		
QY	937	CTATGTGTCCAATGAATGCTCTCTACTGTCAGACCCAGGTCATGCCCTCC--CA	987		
Db	824	-----CysAspThrLeuSer--LysArgSerSerThrSerSerAspHis	837		
QY	988	CCTGCTCTGGCATACCCCATCAAGCAACAAAGAACATTT-----AT	1032		
Db	837	sPheSerAlaSer-----GluCysSerSerGlnGlyGlyPheLysThrLysGlyProLe	855		
QY	1033	CCATACATCTCAATATGCTTCC-----CNAAGT	1059		
Db	855	uHisThrArgGlnCysAsnSerHisSerLysSerAspAsnIleProvalThrProGlnGly	875		
QY	1060	GTGTGCACATGCACGTAAACACACACACACAAATTCAGGTACAGTACGTGGGCAAGT	1119		
Db	875	sCysProSerSerThrGlyPheHisIle-----GlnGluAsnGluGluSe	890		
QY	1120	ATATTCTGCTCATCAATGTCATTGGCTATGTACTTTGTGCAGGGAAGT-----	1169		
Db	890	rHisTyGluSerGlnArgValThrPhe-HisLeuProAspGlySerGlnGluSerC	910		
QY	1170	-----ACATTATCTACAGTCACAAAATGCTCTCATG	1200		
Db	910	ysSerAspSerGlyLeuGlyAspHisGluProValGlySerGlyThrLeuIleSerHisP	930		
QY	1201	GGAAAGCCTTG-----CCAGATTCACACATATATACAAATT-----	1238		
Db	930	roLeuProLeuValGlnProGlnAspGluPheTyAspGlnAlaSerProAspLysArgT	950		
QY	1239	-----CCTAACGACGACGGCCCCCATACACCATCTATTTCAT	1275		
Db	950	hrGluAlaAspGlyAsnSerAspProAsnSerAspGlyProLeuGlyProArg-----	967		
QY	1276	AAACCACTCAGGTTCACAGATGCATGCTTCTCCTATTCTTAACCTACACATAAATTTC	1335		
Db	968	-----GlyLeuAlaGluAlaThrGluMetCysThyThrg	978		
QY	1336	TGGAAGTACTCATATATGGACATTCACAGCACTGCTACAGTCCCACCCCTGTGTGTCT	1395		
Db	978	lngLueCysLeuValLeuGlyHisSerAspAsn-CysTyrMetPro---ProGlyLeuGly	996		
QY	1396	TGATACACAGACACACCAAG	1413		
Db	997	ProTyGlnHisProLys	1002		
RESULT 8					
Q8BIR7					
ID	Q8BIR7	PRELIMINARY;	PRT; 468 AA.		
AC	Q8BIR7;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Weakly similar to zinc finger protein 37A.				
GN	4930488P06RIK.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
NCBI	TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Testis;				
RX	MEDLINE=22354683; PubMed=12466851;				
RA	The FANTOM Consortium,				
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RT	60,770 full-length cDNAs."				
RL	Nature 420:563-573 (2002).				
DR	EMBL; AK029711; BAC26575.1; -.				
DR	MGD; MGI:2442105; 4930488P06RIK.				
DR	InterPro; IPR007087; Znf C2H2.				
DR	InterPro; IPR007086; Znf C2H2.				

Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL: AJ491857; CAD42654.1; -
 DR GO: 0005576; C:extracellular; IEA.
 DR GO: 0005507; F:copper ion binding; IEA.
 DR GO: 0005489; F:electron transporter activity; IEA.
 DR GO: 0005179; F:hormone activity; IEA.
 DR GO: 0007155; P:cell adhesion; IEA.
 DR GO: 0006118; P:electron transport; IEA.
 DR InterPro: IPR000923; BlueCu_1.
 DR InterPro: IPR006207; Cys knot_C.
 DR InterPro: IPR00421; EGF-like.
 DR InterPro: IPR001545; Gly_hormoneB.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR InterPro: IPR002919; TIL_Cysrich.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR008085; TSP_1.
 DR InterPro: IPR006552; VC_out.
 DR InterPro: IPR001007; VWF_C.
 DR InterPro: IPR001846; VWF_D.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR Pfam: PF00057; ldl_recept_a; 10.
 DR Pfam: PF01826; TIL; 10.
 DR Pfam: PF00090; tsp_1; 25.
 DR Pfam: PF00093; vwc; 1.
 DR Pfam: PF00094; vwd; 3.
 DR PRINTS: PRO0261; LDLRECEPTOR.
 DR PRINTS: PRO1705; TSP1REPEAT.
 DR SMART: SM00192; LDLa; 10.
 DR SMART: SM00209; TSP1; 25.
 DR SMART: SM00214; VMC; 4.
 DR SMART: SM00215; VMC_out; 9.
 DR SMART: SM00216; VWD; 2.
 DR PROSITE: PS00196; COPPER_BLUE; 1.
 DR PROSITE: PS01225; CTCK_2; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS00222; FAS8C_3; 1.
 DR PROSITE: PS00261; GLYCO_HORMONE_BETA_1; 2.
 DR PROSITE: PS01209; LDLRA_1; 8.
 DR PROSITE: PS00068; LDLRA_2; 10.
 DR PROSITE: PS00092; TSP1; 25.
 DR PROSITE: PS01208; VWF_1; 1.
 DR PROSITE: PS01184; VWF_2; 2.
 SQ SEQUENCE 4998 AA; 535019 MW; DA2ABA8DA47DF225 CRC64;

Alignment Scores:
 Pred. No.: 0.0249 Length: 4998
 Score: 118.00 Matches: 159
 Percent Similarity: 30.76% Conservative: 64
 Best Local Similarity: 21.93% Mismatches: 155
 Query Match: 1.47% Indels: 347
 DB: 11 Gaps: 45

US-10-776-827-10 (1-4460) x Q8CG65 (1-4998)
 QY 1625 CAAGAGCTCTTTGTCTTGTGGAGA---AAGGCATTTCTGAGTGGTTC---GGG 1572
 Db 689 GlnGluGlyValLeuLeuGlnTrpArgAnGlnThrLeuCysProValSerCysProGly 708
 QY 1571 GGTAAATGTTTCAGAGAGGCTTGGATCATCAATTAACACCTCTCAGTGCAATGAGGAC 1512
 Db 709 GlyGlnValTyGlnGluCys-----AlaProValCysGly----- 720
 QY 1511 GGACTTGTCTAGTGTACAGAAAGAGCTTCAATGAATTTTGAACCATCACACTTTA 1452
 Db 721 -----HisHis----- 722
 QY 1451 AACATCTTGGCAGAGGTGAGGGGTCAGAGGCACAGAACTGTGTCTGTATCAAGAC 1392
 Db 723 -----Cys--GlyGluProGluAspCysGlyGluLeuGlyLe----- 734
 QY 1391 ACACAGGGTGGGACTGTAGCAGGTTC-----TGGAAATGTCC 1353
 Db 1041 nGlyLeuGluProThrAlaTyLeuThrAla---HisSer-TrpArgLeuAsnProL 1060

Db 735 -----CysValAlaGlyCysAsnCysProProGlyLeuLeuTrpAspLeuG 750
 QY 1352 AA---TTATGAGTACTTCCAGTAAAGTTTATGTGAGA----- 1317
 Db 750 ludGlnCysValProProSerMetCysHis-CysGlnPheGlyGlyHisArgTyThr 769
 QY 1316 -----GTTGAAATAGAAAGCATGTCATCTGAACCTGAGTGGTTATGG 1272
 Db 770 IleAsnThrThrThrValArgAspCysSerHisCysileCysGlnGluArgGlyLeuTrp 789
 QY 1271 AAT-----AGATGCTGTATGGGGCCTTCTGCTGTAGGAAA 1236
 Db 790 AsnCysThrAlaHisCysCysProArgGlnTrp-----AlaLeuCysProArgGlu 806
 QY 1235 TTGTATATATGTCTGTAATCTGGCAAGCTTCCCATGAGACATTTTGTGTAGCTAGA 1176
 Db 807 Leu-IleTyVal-----ProGlyAlaCysLeuLeuThrCysAs 819
 QY 1175 TAAT-----GTACTTCCCTG 1161
 Db 819 pSerProArgAlaAsnHisSerCysTrpAlaGlySerThrAspGlyCysValCysProPr 839
 QY 1160 CACAAAGTACATAGCAATGACCAATTTGATGAGCAGCAATATCTCCCACTGCTGCT 1101
 Db 839 oGlyThrValLeuLeuAspLysHisCysValSer----- 850
 QY 1100 ACCTGAATTTGTGTGTGTGTGTACGTGTCACACACACTTTGGG----- 1052
 Db 851 -ProAspLeu-----CysProCysArgHisAsnGlyGlnTrpTyPr 864
 QY 1051 -----ACCATATTGAGATGATGATAAATGTTCTTTGTGCTTTGTCATGGGGT 1002
 Db 864 oProAsnAlaThrIle-----GlnGluAspCysAsnIleCys-----Va 877
 QY 1001 ATGCCAGAGACAGTGGGAGCATGCTGGTCTCGAGTACAGGACATGCTAT---TGG 945
 Db 877 lCysGln-----Gly--GlnArgTrpHisCysThrGlyGlnArgCysSerGlyTrp 893
 QY 944 ACACATAGCTGCCAAGAC----- 927
 Db 894 -----CysGlnAlaSerGlyAlaProHisTyValThrPheAspGlyLeuValPhe 910
 QY 927 ----- 927
 Db 911 ThrPheProGlyAlaCysGluTyLeuLeuValArgGluAlaGlyGlyArgPheSerVal 930
 QY 926 AGTATCCAAAGATTGTCACCGTGGTCCCTGGAGTCCTCCACTGGGACACTTATGCTG 867
 Db 931 SerValGlnAsnLeuProCysGlyAlaSerGlyLeuThrCysThrLysAlaLeuValVal 950
 QY 866 AGGTTTGTCTGTTGAGTAGTTACTCTGTGACATGGACATGACATGCTATTTTCAGG 807
 Db 951 Arg-----LeuAspSerThrValValHisMetLeuArg 961
 QY 806 GGTCTCTGAAATGCTGT-----GTGCTGTCAGAGGA 774
 Db 962 GlyGlnAlaValThrValAsnGlyValSerIleArgLeuProLysValTyThrGlyPro 981
 QY 773 GGCTTAGCCCCAGAAAGAACACCTCTGCTCTGCTGCTAACTAAACAG----- 725
 Db 982 GlyLeuSerLeuHisHisAlaGlyLeuPheLeuLeu-LeuThrThrArgLeuGlyLeuTh 1001
 QY 724 -----GCAGTGATGGTCTCT--TCTCCA----- 704
 Db 1001 rLeuLeuTrpAspGlyClyThrArgValLeuValGlnLeuSerProHisHisGlyAr 1021
 QY 703 -----TCTGAGAAAGCTTGGGTTGAGCAGACATCCAGA-- 671
 Db 1021 gValAlaGlyLeuCysGlyAsnPheAspSerAspAlaSerAsnAspLeuArgSerArgG 1041
 QY 670 -GGTCTCTTCATCT 612
 Db 1041 nGlyLeuGluProThrAlaTyLeuThrAla---HisSer-TrpArgLeuAsnProL 1060


```
QY 897 CAGGACACAGTGTGCAAAATCTTGGTACTGTCTTGGCAGCTATGTGTCCAAATGCAATG 956
Db 138 Phe-----ArgGlnSerAlaLeuProPheHisLeuAlaArgAlaHisProGluIle 155
QY 957 CTCTTACTGACAGCAGCAGTCCCTCCACCTGTCTCTGGCATACCCACATGCAAG 1016
Db 156 IleThrValThrAlaProSer-----ProSerThrLeuThrHisCysThrGlnCysPro 173
QY 1017 CACAAAGACATTATCCATACATCTCAATATGTTGCCAAGTGTGTGCATGCAAGTGA 1076
Db 174 ArgAlaPheHisSerSerAlaGlyLeuArgAsnHisSerArg-----IleHisVal 190
QY 1077 ACACACACACACAAATTCAGGTAGCAGG-----TACGTGGCCAGTAT 1121
Db 191 ValProSerLeuSerAspProGlyThrGluAlaHisLeuCysGlyIleCysGlyLys--- 209
QY 1122 ATTCTCTCATCAATAGGTGCTATGCTGTATCTTTGTGAGGGAAGTACATTATCTACA 1181
Db 210 -----SerPheSerLysSerSerThrLeuThrArg 219
QY 1182 GTCACAAAATGCTCAT-----CGGAAAGCCTTG 1211
Db 220 HisLeuGlnArgHisSerGlyGluLysProPheLysCysProGluCysGlyLysPhe 239
QY 1212 CCAGATTACACACATATATACAAATTCCTAACCAAGGCCCCCATACACCATCTATT 1271
Db 240 LeuGluSerAlaThrLeuValArg-----HisGlnArgThrHisThrGlyGluLys 256
QY 1272 CCATTAACCACTAGGTTCACAGTCATGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1331
Db 257 ProTyAlaCysSerAspCysGlyArgCysPheSerGluSerSerThr----- 272
QY 1332 TTTACTGGAAGTACTCATATATTGACATTCACGACCTCTCAGTCCGCCACCTGTGT 1391
Db 273 -----LeuLeuArgHisGlnArgSerHisGlnGlyGluArgProHisValCys 288
QY 1392 GTCTTG-----ATACAGACACCAAGTCTTGT- 1420
Db 289 AlaThrCysGlyLysGlyPheGlyGlnArgTyAspLeuValHisGlnArgSerHis 308
QY 1421 ---GCCTCTGACCTCCTACCTGCGCAAGATGTTTAAAGTGTGTGTTTCAAAATTCATT 1477
Db 309 ThrGlyGluArgProPheProCysProGlnCys-----GlyArgGlyPheThr 324
QY 1478 GAAAGCTCTTTCTTGTA-----ACTCATGACAAAGTCCTCT---CATTCG 1522
Db 325 AspArgSerAspLeuThrLysHisLeuArgThrHisThrGlyGluLysProTyHisCys 344
QY 1523 CAC-----TGAGAGTGTGTTAATGTGATCCAGACCTCTCTGTGAAACATTTACCCCGC 1576
Db 345 GluLeuCysGlyLysArgPheThrCysIleSerAsnLeuAsnVal-----HisLeuArg 362
QY 1577 AAACCACTCAGCA-----AAGTGCCTTTCTTCCAGCAAGAAACAAAGCTCTT 1624
Db 363 -AsnHisAlaGlyHisLysProHisLysCys-----ProGluCysGlyLysSerPheSe 380
QY 1625 GGTGTGTACTCTAGAAAATATGAAGCCCTCATTTATGTCTAGTGCAGTCAACTGT 1684
Db 380 rValalaSerLysLeuAlaLeuHisArgLysThrHisLeuGlyGluArgThrAlaGlu-C 400
QY 1685 GTACC---TGTGCAATGTTTACAGATGGAAGGGTGAGGAGATGTCTACACCTGACGTAG 1741
Db 400 ysThrGluCysGlyLysPhe----- 406
QY 1742 TATCTCTATATAACCAAGTTTCCAGCAGGGAAGGAATAGACAATCATCAGTGCAGTC 1801
Db 407 -----PheSerHisGlyArgSerLeuSerGlnHisGln--ArgSe 419
QY 1802 TCACAGAGG 1811
Db 419 rHisArgArg 422
```

RESULT 11

```
Q8CJ95
ID Q8CJ95 PRELIMINARY; PRT; 1381 AA.
AC Q8CJ95;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Tensin 2.
GN TENC1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lo S.H.;
RT "Molecular cloning and characterization of mouse tensin 2.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF424789; AAN32753.1; -.
DR MGD; MGI:2387586; Tenc1.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR008973; C2 CALB.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR006020; PTB_PID.
DR InterPro; IPR000980; SH2.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00017; SH2; 1.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00109; C1; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50001; SH2; 1.
SQ SEQUENCE 1381 AA; 149969 MW; D1C41CA0F0017786 CRC64;
```

Alignment Scores:

Pred. No.:	Score:	Length:
0.0375	114.50	1381
Percent Similarity:	33.26%	Matches: 104
Best Local Similarity:	22.46%	Conservative: 50
Query Match:	1.43%	Mismatches: 147
DB:	11	Indels: 162
		Gaps: 24

US-10-776-827-10 (1-4460) x Q8CJ95 (1-1381)

QY 2513	TGGTACTACCTCCCTCCAGTGTTACTTAGGCTGTGGAGGGTGGCATTTAGTGTG	2572
Db 694	TrpSerSerGluValGlyLysProLeuLeuHisProValArgProGlyHisProLeuPro	713
QY 2573	ACCTTGACACGAGGGTCTTTCTAACAGATGACCTGTGAATCATATAATTTAAACCTGCATAT	2632
Db 714	LeuLeuValProAla-----Cys-GlyHisHis-----	722
QY 2633	ATTATTAGCAGCTCACATTTGCCCTCTCACCTATATGGC---CATAACTGCCTAAG-	2688
Db 723	-----His-AlaProMetProAspTyArgCysLeuLysProProLysV	737
QY 2689	-----CACTCAGCGCTCCCACTCATCAACCCCTTGCACAGAGAAGAGC	2734
Db 737	alGlyGluGluGlyHisGlyCysSerTyAlaValCysSerGluGlyArgTyGlyH	757
QY 2735	ACTCTGGTTCTTCTATCCCTTGTCCATAGAGAGTTTGTATGGGCTCTGG-----	2787
Db 757	isSerGly-----TyrProAlaLeuValThrTyGlyTyGlyGlyAlaV	772
QY 2788	-----CTGTGCCCTTCACATAACAGAATA-----ACTTGCCATCTGCCTGCA	2829
Db 772	alProSerTyCysProAlaTyGlyArgAlaProHisSerCysGlySerProSerGluG	792
QY 2830	-----CCAAACCCAGGATGTGGAGACATCTCCCAACAACTGCCACTGCTCAC	2878
Db 792	lyArgGlyTyProSerProGly-----AlaHisSerProArgAlaGlySerValSerP	810

QY 2879 CAGACAAA----- 2886
 DB 810 roGlySerProProTyrLeuGlnProArgLysLeuGlyTyrGluIleSerAlaGluAspG 830
 QY 2887 -----GTCGCCCTTCCCTCTCCACCTCTC 2911
 DB 830 lyArgAspLysTyrProLeuSerGlyHisLeuAlaSerThrGlyProLeuAlaSerThrG 850
 QY 2912 AGTCCCTAGATGAGGTG---GGGAGAGGTGGAGGTGACAGCTGAGACGTAGTG 2968
 DB 850 luSerProGluProSerTyrArgAspGlySerSerGlyHisSerThrLeuProArgSerP 870
 QY 2969 TCAGA-----TATGATCTAGGAGGGGATCATCCGGATCCCGG 3007
 DB 870 roArgAspProGlnCysSerAlaSerSerGlyProSerThrProLeu----- 888
 QY 3008 ACCATACAGTAACATGTTTCCATGGCAACTGCTTGTCTGTTGAATTAAGACAGCAGT 3067
 DB 889 --HisThrSerSerProValGlnGlyLysGluSerAsnArgGlnAspThrThrArgS 908
 QY 3068 CAGTTGTCTATTCATGACAGAGGCTCTATCTCCAGGCACAATGCCCTGTCTCTCTA 3127
 DB 908 erProSerLeuAlaProThrGlnArgLeuSerProGlyGlu-----AlaLeuProS 925
 QY 3128 ATCCAAATGACTTGTCTCACCCAGGAGTGAACACCCAGAACTCCTTCTCAGTCAC 3187
 DB 925 erValVal-----GlnGly-----ValAlaGluLysT 934
 QY 3188 TTCACAGCGATGACTCAGAGAGCAAAACCAGAAATGGGCTCTCTTTTCCCATCA 3247
 DB 934 hrProGluLeuLeuThrSerSerArgProGluGlnLeuAspProSerProPheSerGlnT 954
 QY 3248 CAGACTCCCTGACAACTTTTC---TGGCGTAACCTAGAGGAGTCCAGCTGAGGATAGG 3304
 DB 954 hrSerAlaProGlySerProAsnGlyTyrProGlnGluArgSerProGlyGly----- 971
 QY 3305 CCCTAAACGTTTGTAAATAAACAGGTGCATGAAGAGGCCTAAGGCCATTGTTGATAT 3364
 DB 972 -----HisThrAsnSerAlaSerProArg----- 979
 QY 3365 CCACCTCTCTTTCCT 3415
 DB 980 -----SerProValProThrThrLeu-ProGlyLeuArgHisAlaProTyrGln 995
 QY 3416 -----TCTCTGATTCCT 3457
 DB 996 GlyProArgGlyThrSerAspSerProAspGlySerProLeuThrProValProThrGln 1015
 QY 3458 -----CTTTATTCCT 3508
 DB 1016 MetProThrLeuValGlySerProGluProProGlnSerSerProThrProAlaPhePro 1035
 QY 3509 CTGGCAACC---CATGGGACTCAGGACCCAGAGACTCTTCTCTCTCTCTCTCTCTCTCT 3565
 DB 1036 LeuAlaThrSerTyrAspAlaAsnGlyPro----- 1045
 QY 3566 GTTCACGGGGGACAAAAATGATTCCTAAAGAGAGGCTTCTCTAGACACAGCAGCTC 3625
 DB 1046 -----Ile 1046
 QY 3626 CAGAAAGACATCCCTAGGCTGAGCTTCTGAGCAGCTTTTAGCCAGGCTCCGGAGCGAG 3685
 DB 1047 GlnProLeuPro-----GluLysArgHisLeuProGlySerGlyGlnGln 1062
 QY 3686 CCA 3688
 DB 1063 Pro 1063
 RESULT 12
 Q8CGB6 PRELIMINARY; PRT: 1400 AA.
 ID Q8CGB6
 AC Q8CGB6
 DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE Tensin 2.
 GN TENC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;
 RN [1]
 SEQUENCE FROM N.A.
 RP STRAIN=FVB/N; TISSUE=Salivary gland;
 RC Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC042190; AAH42190.1; -.
 DR MGD; MGI:2387586; Tenc1.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR008973; C2_CaLB.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR006020; PTE_PID.
 DR InterPro; IPR000980; SH2.
 DR Pfam; PF00130; DAG_PE-bind; 1.
 DR Pfam; PF00017; SH2; 1.
 DR ProDom; PD000093; SH2; 1.
 DR SMART; SM00109; Cl; 1.
 DR SMART; SM00462; PTB; 1.
 DR SMART; SM00252; SH2; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
 DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
 DR PROSITE; PS50001; SH2; 1.
 SQ SEQUENCE 1400 AA; 152012 MW; 72CB109A17BB021C CRC64;
 Alignment Scores:
 Pred. No.: 0.0377 Length: 1400
 Score: 114.50 Matches: 104
 Percent Similarity: 33.26% Conservative: 50
 Best Local Similarity: 22.46% Mismatches: 147
 Query Match: 1.43% Indels: 162
 DB: 11 Gaps: 24
 US-10-776-827-10 (1-4460) x Q8CGB6 (1-1400)
 QY 2513 TGTACTCTACCTGCCCTCCCTCCAGGTGACTTTAGGCTGTGGAGGTGGCATTTAGTGTG 2572
 DB 717 TtpSerSerGluValGlyLysProLeuLeuHisProValArgProGlyHisProLeuPro 736
 QY 2573 ACCCTGCGACGAGGTTTCTAACAGATGACCCCTGTGAATCATAATTTAAACCTGCATAT 2632
 DB 737 LeuLeuValProLa-----Cys-GlyHisHis----- 745
 QY 2633 ATTTTATAGCAGTACATTTGCCCTCTACCTATATATGGC---CATAACTGCCTAAG- 2688
 DB 746 -----His-AlaProMetProAspTyrGlyCysLeuLysProProlysV 760
 QY 2689 -----CACTAGGCTCCCATCATCAACCCCTTTGACACAGAAAGAAC 2734
 DB 760 alGlyGluGluGlyHisGluGlyCysSerTyrAlaValCysSerSerGlyTyrGlyH 780
 QY 2735 ACTCTGTTCTCTATCCCTTGTACATAGAGTTTGTCTATGGGCTCTGG----- 2787
 DB 780 isSerGly-----TyrProAlaLeuValThrTyrGlyTyrGlyAlaVal 795
 QY 2788 -----CTGTGCCCTTTCATACAGAATA-----ACTTGCATCTCGCTGCA---- 2829
 DB 795 alProSerTyrCysProAlaTyrGlyArgAlaProHisSerCysGlySerProSerGluG 815
 QY 2830 -----CGAACCCGAGGATGTGGAAGACATCTCCCACTCCCACTGCTCTCAC 2878
 DB 815 lyArgGlyTyrProSerProGly-----AlaHisSerProArgAlaGlySerValSerP 833
 QY 2879 CAGGACAA----- 2886
 DB 833 roGlySerProProTyrLeuGlnProArgLysLeuGlyTyrGluIleSerAlaGluAspG 853

Db	372	TrpSerProCys	:	---	375
QY	769	TAGCCCCAGAAAGAAACACCTCTGCTTCCTGCCTAACTAAACAGACGACGTGATGCTCT	:	---	710
Db	376	-----SerGlyThrCys	:	-----Thr	380
QY	709	TCTCCATCTGGAGAGCTGGTTCAGCAGAGATCCAGAGGTCTTCTTCATCCTTCTCT	:	-----	650
Db	381	AspProAlaHisProAlaTrp	:	-----ArgSerArgSerArgLeuCysLeuAlaAsn	397
QY	649	TCTTCTCTCTCTTTTCTCTCT	:	-----	629
Db	398	CysThrGlyGlyAlaAlaSerGlnGluArgProCysAsnLeuProSerCysThrGluLeu	:	-----	417
QY	628	---CTGGCTGTGGCCCTCTTTCCTGTGGTGGCTCCAGCAGC	:	-----	590
Db	418	ProLeuCysProGlyProGlyCysGluAlaGlyAsnCysSerTrpThrAlaTrpAlaPro	:	-----	437
QY	589	-----AATGCCACAGATGAAGAGAGCTGCATCATCACCAGAGAGTGGCAGGC	:	-----	542
Db	438	TrpGluProCysSerArgSerCysGlyValGlyGlnArgArgLeuArgAlaTrpHis	:	-----	457
QY	541	CCAGGAGGACTCCAGGCAGATGGTGTGTAGATGCCTTCTT	:	-----	500
Db	458	ProProGlyProGlyGlyHisTrpCysProAspValLeuThrAlaTrpGlnGluArgArg	:	-----	477
QY	499	-----GGTGTTCGTCAAGATGGTGTCCAGATAG	:	-----	470
Db	478	PheCysAsnLeuArgAlaCysProValProGlyGlyTrpSerArgTrp	:	-----	493
QY	469	AGCTGAAGAGCCACTGATTCCTCCATGGCTGGCTAGTATTAGGCCAGAGAACCT	:	-----	410
Db	494	-----SerProTrp-SerTrpCysAsp	:	-----ArgSerCys	503
QY	409	CTGGCTTACAAGCAGTCTCGAGCTCAAAAGGATGCAACCTTGACCACCTTGTAATTAT	:	-----	350
Db	503	sGlyGlyArgSer---LeuArgSerArgSerCys	:	-----	514
QY	349	TCCTCTCCAGCTTTGGGTTTGGCATGCCACTCACAGCACCTCCACACACAGCAGCAA	:	-----	290
Db	515	-----SerSerPro-ProProLysAsnG	:	-----	522
QY	289	ATGGCCAGCAGCTTGGGAGCCTCTTCCACGTTACTTCTTGCTTGGTGTGTCAGGG	:	-----	230
Db	522	lyGlyAlaProCysValGlyGluArgHisAlaArgLeuCysAsnProThrProCysG	:	-----	542
QY	229	ATCTCGGCCAGACTGCTTTCCTGGAGTACGTTCAACGTAGTAAACAATTCGTGAACA	:	-----	170
Db	542	leuGluGly---Cys---ProAlaGly	:	-----	548
QY	169	CCGCGCCTGTGCATTAGATCTCATCCCGCCAGTTCAGTGTGTCTTGGGAAGGT	:	-----	110
Db	549	--MetGluValValSerCysAlaAsnArgCysProArgArgCysSerAspLeuGlnGluG	:	-----	568
QY	109	CACTCAGCTGCTTTGAGTCTCAGTTTCTTCATCTCCAAATGGG	:	-----GAAC	62
Db	568	lylleValCysGlnGluAspGln-----AlaCysGlnGlnGlyCysArgCysProGluG	:	-----	586
QY	61	AATCTTTTCGTCCGCGGAAGGGCTTCCTG	:	-----CAGGCC	26
Db	586	lySerLeuGlnAspGlyGlyCysValProLeuGlyHisCysGluCysThrAspAla	:	-----	606
QY	25	AGGGAATCTCTGG	:	-----	12
Db	606	lnGlyHisSerTrp	:	-----	610

QY	229	AATCGGCCAGAACTGCTTTTCCTGGAGGTCAACGTTTCAAAGTAACAATAA	CCTCCTGAACA	170
Df	:			
Df	4051	luGlugly-----Cys---ProAlagly-	-	4057
QY	169	CCGCCTTGCAATTGATCATCACGCCAGTTCAGTGTGTCTCTTGGGAAGGT		110
Df	:			
Df	4058	--MetGluValSerCysAlaAsnArgCysProArgCysSerAspLeuGlnGlu	G	4077
QY	109	CACTCAGCTCTTTGAGTCTCAGTTTTCTTCATCTCCAATAATGG--	-GAAC	62
Df	:			
Df	4077	lyrleValCysGlnGluaspGln----AlaCysGlnGlnGlyCysArgCysProGlu	G	4095
QY	61	AATCTTCGGTCCGCGAAGGGTCTCTG----	-CAGGCC	26
Df	:			
Df	4095	lyrSerLeuGlnAspGlyGlyCysValProLeuGlyHisCysGluCysThrAspAla	G	4115
QY	25	AGGAATCTCCTGG	12	
Df	:			
Df	4115	InGlyHisSerTrp	4119	

RESULT 15
Q7Z5J3 PRELIMINARY; PRT; 451 AA.

ID	Q7Z5J3	PRT;	451 AA.
AC	Q7Z5J3;		
DT	01-OCT-2003 (TrEMBLrel. 25, Created)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	BNFL		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=12853144;		
RA	Wu I.; Moses M.A.;		
RT	"BNF-1, a novel gene encoding a putative extracellular matrix protein,		
RL	is overexpressed in tumor tissues."		
RL	Gene 311:105-110(2003).		
DR	EMBL; AV163868; AAO31809.1; --		
SQ	SEQUENCE 451 AA; 49675 MW; CPBF0BFA7D58CFE3 CRC64;		

Alignment Scores:	
Pred. No. :	0.0326
Score:	Length: 451
Percent Similarity:	Matches: 90
Best Local Similarity:	Conservative: 38
Query Match:	Mismatches: 124
DB:	Indels: 134
	Gaps: 23

US-10-776-827-10 (1-4460) x Q7Z5J3 (1-451)

QY	318	GTGGCATGCCAACCCAAAGCTGGAAGGAATAAATTACAAGTGTCAAGTTCATC	377
Df	:		
Df	79	ValHisCysProGlnPro-----ValThrGluProGlnGlnCysCys	92
QY	378	CTTTTGAGCTCAGGACCCTGTGT---	422
Df	:		
Df	93	-----ProLysCysValGluProHisThrProSerGlyLeuArgAlaPro	107
QY	423	CCTAATCTAGCCAAGCAC-----CATGGAGAG-----	449
Df	:		
Df	108	ProLysSerCysGlnHisAsnGlyThrMettyrGlnHisGlyGluIlePheSerAlaHis	127
QY	450	-----AATCAGTSCCTTCTT-----	467
Df	:		
Df	128	GluLeuPheProSerArgLeuProAsnGlnCysValLeuCysSerCystThrGluGln	147
QY	468	CTCTATCTGGGACCACTCTTGACCACAAAACCAAGAAGGCATCTACACCACTGTGCT	527
Df	:		
Df	148	IleTyrcysgly-----LeuThrThrCysProGlu-----	158
QY	528	GGGAGTCTCCTGGGCGCTCCACTTTGGTGATCATCACAATCTCTTTCATCTGTGCA	587

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:07:49 ; Search time 127 Seconds
(without alignments)
6756.129 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaaggtactactagctc 4460

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p_model -DEV=xlp
-Q=/cgn2_1/USFTW_spool_p/US10776827/runat_21092004_133241_20790/app_query.fasta_1.4615
-DB=PIR_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISF=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1 221 @runat_21092004_133241_20790 -NCPU=6 -ICPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	1.5	319	1 RWRBB	T-cell receptor be
2	117.5	1.5	628	2 S01955	hypothetical prote
3	113	1.4	628	2 S19150	hypothetical prote
4	112.5	1.4	631	1 A36749	transcription fact
5	112	1.4	597	2 JQ0107	hypothetical 66K p
6	109	1.4	968	2 T45746	hypothetical prote
7	108	1.3	322	2 S05500	Ig alpha-1 chain C
8	107	1.3	628	2 JQ0110	hypothetical 69K p
9	107	1.3	3020	2 A43932	mucin 2 precursor,
10	106.5	1.3	2843	1 RBHUAP	adenomatous polypo
11	105	1.3	630	1 A46149	transcription fact
12	104	1.3	1184	2 G01763	atrophin-1 - huma
13	103.5	1.3	330	2 S37595	mucin JUL10 - huma
14	103	1.3	353	1 A1HU	Ig alpha-1 chain C

15	102.5	1.3	649	2 T46500	hypothetical prote
16	102	1.3	962	2 S03818	carboxymethylcellu
17	101.5	1.3	628	1 A33333	transcription fact
18	101.5	1.3	769	1 JC1121	leukocyte adhesion
19	101.5	1.3	1184	2 S50832	atrophin-1 - human
20	101	1.3	1959	1 AGRT	agrin - rat
21	101	1.3	1964	2 T09059	otogelin - mouse
22	101	1.3	2910	2 T42214	otogelin - mouse
23	100.5	1.3	1210	2 I39410	AF-4 protein, spli
24	100.5	1.3	2825	2 T14271	Doc4 protein, stre
25	100	1.2	992	2 T38817	hypothetical prote
26	100	1.2	1257	2 S28764	neurocan precursor
27	100	1.2	3507	2 T34513	hypothetical prote
28	99.5	1.2	660	1 Q0BE3	BHLF1 protein - hu
29	99	1.2	140	2 T46880	T-cell receptor be
30	99	1.2	753	2 JQ0532	OP protein - Kenne
31	99	1.2	2453	2 S60254	nuclear receptor c
32	98.5	1.2	542	2 S39608	transcription fact
33	98.5	1.2	543	2 S35047	mucin JUL7 - human
34	98.5	1.2	2395	1 S50820	surface protein ty
35	98.5	1.2	2845	2 T49505	adenomatous polypo
36	98	1.2	589	2 B38128	epithelin/granulin
37	98	1.2	814	1 I39627	nicotine dehydroge
38	97.5	1.2	1042	2 T48801	hypothetical prote
39	97.5	1.2	1213	2 A58198	serine/proline-ric
40	96.5	1.2	344	1 RWRTC2	T-cell surface gly
41	96.5	1.2	557	2 A55933	paxillin - human
42	96.5	1.2	628	1 A39282	transcription fact
43	96.5	1.2	1548	2 S34583	serine proteinase
44	96.5	1.2	1762	2 T03222	probable polykeid
45	96	1.2	333	2 G96780	unknown protein P9

ALIGNMENTS

RESULT 1

RWRBB
T-cell receptor beta chain precursor (ANA 11) - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 30-Jun-1987 #sequence revision 30-Jun-1987 #text_change 16-Aug-1996
C:Accession: A94052; A94616; A02136
R:Angiolillo, A.L.; Lamoyi, E.; Bernstein, K.E.; Mage, R.G.
Proc. Natl. Acad. Sci. U.S.A. 82, 4498-4502, 1985
A:Title: Identification of genes for the constant region of rabbit T-cell receptor beta
A:Reference number: A94052; MUID:85242712; PMID:2989826
A:Accession: A94052
A:Molecule type: mRNA
A:Residues: 1-319 <ANG>
R:Mage, R.
submitted to the Protein Sequence Database, April 1987
A:Reference number: A94616
A:Contents: corrections to residues 79-83 and 104-105
A:Accession: A94616
A:Molecule type: mRNA
A:Residues: 79-83,104-105 <MAG>
C:Superfamily: immunoglobulin C region; immunoglobulin homology
C:Keywords: glycoprotein; heterotetramer; T-cell receptor; transmembrane protein
F:166-240/Domain: immunoglobulin homology <IMM>
F:292-313/Domain: transmembrane #status predicted <TM>
F:314-318/Domain: intracellular #status predicted <INT>
F:147,164,254,262/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:			
Pred. No.:	0.0151	Length:	319
Score:	123.50	Matches:	78
Percent Similarity:	34.22%	Conservative:	38
Best Local Similarity:	23.01%	Mismatches:	112
Query Match:	1.54%	Indels:	111
DB:	1	Gaps:	15

US-10-776-827-10 (1-4460) x RWRBB (1-319)

QY 954 ATGCTCTTACTGACAGCCAGCATGCTCCACCTGT----- 992

QY	3276	TAACTAGAGGATCCCACTGCAGTACGAGTAAGCGCCTAAACAGTTTGTAAATAAACAAGTGCA	3333
Db	275	uProArgThrSerProArgGly-----LeuLeuProAsnProArgArgHi	291
QY	3336	TGAAGAAGGACTAAGGCCATTGTGATATCCACTCTCTTTCCCACCTTCCTTCTCATCT	3395
Db	291	sArgThrSerThr-GlyHis-----	297
QY	3396	TTTTTCCATGTTTTTATGCTTCTCTGATNTCCCTCTTCTGCTGCACGACCAGCC	3451
Db	298	-----lIeProProThrThrSerArgProThrGlyP	309
QY	3452	--CCAGCCCTTTATCTCTCTCCATTTTCACCTCTCCAGCTCTGTCCTGAACTGCCAC	3509
Db	309	roProSerArgLeuGlnArgPro-ValHiLeuTyrgLnSer--SerProHISThrProAs	328
QY	3510	TGGCAACCCATGGGACCTCAGGACCAGACTGCTTCACTCATCTGGGG-----	3558
Db	328	nPheArgProSerSerllearg---LysASPAlaleuLeuGlnThrGlyProArgLeuGl	347
QY	3559	-----AGGTAAGTTTCACCGGGGACAAAAA	3584
Db	347	yHisLeuGluArgLeuGlyGlnProAlaAsnLeuArgThrSerGluArgSerProPro--	366
QY	3585	ATGATCTCTAAAGAGAGGCTTCCTAGACAGCACAGGCTCCAGAAAGACATCCCTTAGG	3644
Db	367	-----ThLyArgArgLeuProArgSerSerGlu---ProasnArg--LeuProLys	382
QY	3645	CCTGGACTTCTGACGACGCTTAGCCAGGCTCCGAGCGGACGCCAGAGAGGCGCTTCCCC	3704
Db	383	Pro-----LeuProGluAlaThrLeuAlaProSerTyrgArgHisA-gA-gProTygPro	400
QY	3705	ATTGCTCTCTTCCCATTGCTCAATGATTCCARGTTTCTTTTCTGGGGGAGACAGGG	3764
Db	401	LeuLeuProAsnProProAlaAla---LeuProSerileAlatyThrSerSerArgGly	419
QY	3765	AGG 3767	
Db	420	Lys 420	
RESULT 3			
S19150 hypothetical protein, 69k - turnip yellow mosaic virus			
C;Species: turnip yellow mosaic virus, TYMV			
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-2000			
C;Accession: S19150			
R;Dreher, T.W.; Bransom, K.L. Plant Mol. Biol. 18, 403-406, 1992			
A;Title: Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a CDNN			
A;Reference number: S19150; MUID:92119261; PMID:1731998			
A;Accession: S19150			
A>Status: preliminary			
A:Molecule type: genomic RNA			
A;Residues: 1-628 <DRE>			
A;Cross-references: EMBL:X16378; NID:g62218; PIDN:CAA34414.1; PID:g62219			
C;Superfamily: hydroxyproline-rich glycoprotein			
Alignment Scores:			
Pred. NO.:	0.154	Length:	628
Score:	113.00	Matches:	103
Percent Similarity:	31.53%	Conservative:	37
Best Local Similarity:	23.20%	Mismatches:	139
Query Match:	1.41%	Indels:	165
DB:	2	Gaps:	22
US-10-776-827-10 (1-4460) x S19150 (1-628)			
QY	2748	ATCCCTTGTCACATAGAGAGTTGTGTCATGGGCGCT-----CTGGCTGTGCCCTTC	2798
Db	40	LeuProMetValHisSerGlnGlyThrSerThrProThrGlnLeuLeuArgHisProAsn	59
QY	2799	ACATAACAGAAATACTTGCCATCTGCTTGCACCTGCAACCAACCCAGGATGTGGAAGACATCTCC	2858

Db	60	IleTrpPheGlyAsnLeuProProProArgArgProGlnAspAsnArgAspPheSer	79
Qy	2859	CCACAACTGCACCTGCTC	2876
Db	80	ProLeuHisProLeuValPheProGlyVhiHisSerGlnLeuArgHisValHisGluThr	99
Qy	2877	-----ACCAGGACAAGCTCCCTTCCTGCTCCACCTCTCAGTCCCCCTAGAAATGGATG	2930
Db	100	GlnGlnValGlnThrCysProGlyLysLeuLysLeuSer	113
Qy	2931	GCTGGGAGAGGTGGAGGCTGACAGCTGACAGCTAGTGTGATGATCTAGGAGGCG	2990
Db	114	-----GlyAla115	
Qy	2991	GATCACCGGATCGGACCATACAAGTAAATCGT	3044
Db	116	GluGluLeuProAlaProGlnArgGlnHisSerLeuProLeuHis-IleThrArgPro	135
Qy	3045	CTCG-----TTTGAATTAAAGACAGCAGCTCAGTT	3072
Db	135	oSerArgPheProHisHisPheHisAlaArgProAspValLeuProSerValProAs	155
Qy	3073	-----GTCAATGCCATGACAAGGCTCTACT	3099
Db	155	pHisGlyProValLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerAl	175
Qy	3100	-----CCAGGCAATGTCCCTGCTGCTCCCTAAAT	3129
Db	175	aThrArgGlyProSerPheArgProLysLeuLeuProLysValValHisValHisAspAs	195
Qy	3130	-----CCAATGGACTGTCTCACCCAGGGATGAACACCCAGAACTCACT	3177
Db	195	pProProHisSerLeuArgProArgGlySerArgSerArgGlnLeuGlnProThrVa	215
Qy	3178	-----TCTCAGTCACCTCCACAGCCGATGAC	3203
Db	215	lArgArgProLeuLeuAlaProAsnGlnPheHisSerProArgGlnProProLeuSe	235
Qy	3204	TCAGAGAGCCAAACCCAGAATGGGCTCTCTTTTC	3240
Db	235	rAspAspProGlyLeuGlyProArgProLeuAlaProAsnSerThrArgAspProPr	255
Qy	3241	-----CCCATCACAGACTCCCCTGCACAAC	3275
Db	255	oProArgProIleThrProGlyProSerAsnThrHisGlyLeuArgProLeuSerVal	275
Qy	3276	TAACTAGAGAGTCCAGTGCAGATAGGCGCTTAACGTTTGTAAATAAACAGGTGCA	3335
Db	275	uProArgAlaSerProArgGly	291
Qy	3336	TGAAGAGGCTAAGGCCATTGTTGATATCCACTCTCTCTTCCACTT	3387
Db	291	sArgThrSerThrGlyHisIleProProThrThrThrSerArgProThrGlyProProSe	311
Qy	3390	TCATCTTTTTCCTCCATGTTTATGCTCTCTGATTCCCTCTCTGCTGTCACACGACAG	3445
Db	311	rArgLeuGlnArgProVal	323
Qy	3450	CCCCAGCGCTTTATTCCTCTCCATTTTCACTCTCTTCCAGCTC	3492
Db	323	rPro-----HisThrProAsnPheArgProSerSerIleArgLysAspAlaLeuLe	340
Qy	3493	-----TGTCCTCGAATGCCACTGCCACTGCCACCCCATG	3521
Db	340	uGlnThrGlyProArgLeuGlyHisLeuGluCys	356
Qy	3522	GGACCTCAGACACAGAGACTGCTTGACTCATCTGGGAGGGTAAGTTTACCGGGGACAAA	3581
Db	356	aAsnLeuArgThr	366
Qy	3582	AAATATGATCTTAAAGAAGAGGCTTCTTAGACACGACAGGCTCCAGAAAGACATTC	3641

Db 366 o-----ThrLysArgLeuProArgSerSerGlu---ProAsnArg--LeuPro 381
 QY 3642 AGCCCTGGACTCTGACGAGCTTTAGCCAGGCTCCGGACGGCAGCAGAGGAGGCTTTC 3701
 Db 382 LysPro-----LeuProGluAlaThrLeuAlaProSerTyrArgHisArgArgSerTyr 399
 QY 3702 CCATTGCTCTTCCCATTCCTCAATGATTCCCATGTTCTTTTCTTGGGGGAGCA 3761
 Db 400 ProPheLeuProAsnProProAlaAla---LeuProSerIleAlaTyrThrSerSerArg 418
 QY 3762 GGGAGG 3767
 Db 419 GlyLys 420
 RESULT 4
 A36749
 N;Alternate names: hepatic nuclear factor 1-alpha; transcription factor 1, hepatic; tran
 C;Species: Homo sapiens (man)
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: A36749
 R;Bach, I.; Galcheva-Gargova, Z.; Mattei, M.G.; Simon-Chazottes, D.; Guenet, J.L.; Cere
 Genomics 8; 155-164, 1990
 A;Title: Cloning of human hepatic nuclear factor 1 (HNF1) and chromosomal localization o
 A;Reference number: A36749; MUID:91184801; PMID:1707031
 A;Accession: A36749
 A;Molecule type: mRNA
 A;Residues: 1-631 <BAC>
 A;Cross-references: GB:M57732; NID:g184264; PIDN:AAA88077.1; PID:g184265
 C;Genetics:
 A;Gene: GDB:TCF1; HNF1; LFB1
 A;Cross-references: GDB:125297; OMIM:142410; OMIM:600496
 A;Map position: 12q24.3-12q24.3
 C;Complex: homodimer; can also form heterodimers with, for example, HNF-1B
 C;Function:
 A;Description: transcription activator required for the expression of a number of liver-
 A;Note: also expressed in some other tissues, where it may play other roles
 C;Superfamily: transcription factor HNF-1; homeobox homology
 C;Keywords: DNA binding; heterodimer; homeobox; homodimer; liver; nucleus; transcription
 F;1-33/Region: dimerization
 F;222-277/Domain: homeobox homology <Hox>

Alignment Scores:

Pred No.: 0.172 Length: 631
 Score: 112.50 Matches: 103
 Percent Similarity: 31.55% Conservative: 44
 Best Local Similarity: 22.10% Mismatches: 180
 Query Match: 1.40% Indels: 139
 DB: 1 Gaps: 23

US-10-776-827-10 (1-4460) x A36749 (1-631)

QY 225 GAGATCCCTGACACGCAAGCAAGAGTAACTGGAGGAGGCTCCCAAGCTGGCTG 284
 Db 193 GluLeuPro-----ThrLysLysGlyArgArgAsnArgPheLysTrpGlyProAlaSerGln 211
 QY 285 GCCATTTCGTGCTGTGTGTGGAGGTGCTGTCAGTGGCATGCCCAACCCAAAGCTGAA 344
 Db 212 GlnIleLeuPheGlnAlaTyrGlu-----ArgGlnLysAsnProSerLysGlu 227
 QY 345 GAGGAATAAATACAAAGTGGTGCATTCCTCTTTTGACTCAGGACCTGCTGTGAAG 404
 Db 228 GluArg--GluThrLeuValGluGluCys-----AsnArgAlaGluCysIle 242
 QY 405 CCGAGAGGGTTCCTGGCCCTTAATCTAGCCAAAGCACCAGCATGAGAGATCACTGCTCTT 464
 Db 243 GlnArgGlyValSer--ProSerGlnAlaGlnGlyLeuGlySerAsnLeuValThrGlu 261
 QY 465 CAGCTCTATCTGGGACACCATCTTGACCAACACCAAGAGGCATCTACACACCATCTG 524
 Db 262 ValArgValTyrAsnTrpPheAlaAsnArgArgLysGluGluAlaPheArgHisLysLeu 281
 QY 525 -----CCTGGAGTCTCTCTGGCCCTGCCACTCTTGGT 557

Db 282 AlaMetAspThrTyrSerGlyProProGlyProGlyProGlyProGlyProAla---LeuPro 300
 QY 558 GATCATCACACTCTCTTTCATCTGTGGTCCATCTGCTGGAGCCACCAGGAGGGGG 617
 Db 301 AlaHisSerSerPro---GlyLeuProProAlaLeuSerProSerLysValHisGly 319
 QY 618 -----CCAGCAGCCAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAGATGA 662
 Db 320 ValArgTyrGlyGlnProAlaThrSerGluThrAlaGluValProSerSerSerGly--- 338
 QY 663 AGAAGACCTCTGGATCTCTGCTCAACC-----CAAGTCTCTCAAGATGAGAGAG 713
 Db 339 ---GlyProLeuValThrValSerThrProLeuHisGlnValSerProThrGlyLeuGlu 357
 QY 714 ACCATCAGCTGCTGTTAGTTAGGAGGAGGAGAGGTGTTCTTCTGGGCTAAGCC 773
 Db 358 Pro---SerHisSerLeuLeuSerThrGluAlaLysLeuValSerAlaAlaGly----- 374
 QY 774 TCCTCTGACCCACACAGACATTTTCAGGAACCCCTGAAATATGCACTATGTCCATGTC 833
 Db 375 -----GlyProLeuProProValSerThrLeuThrAlaLeu 386
 QY 834 CACAGA-----GTAACACTCAACCAAGAGAAACAACTCAGACT 872
 Db 387 HisSerLeuGlnThrSerProGlyLeuAsnGlnGlnProGlnAsnLeuIleMetAla 406
 QY 873 AAGTGTCCAGTGGAGGAGGAGTCCAGGAGCA----- 905
 Db 407 SerLeuPro---GlyValMetThrIleGlyProGlyGluProAlaSerLeuGlyProThr 425
 QY 905 ----- 905
 Db 426 PheThrAsnThrGlyAlaSerThrLeuValIleGlyLeuAlaSerThrGlnAlaGlnSer 445
 QY 906 -----CTGGACAATTCCTGGATACCTCTCTGGAGCT----- 938
 Db 446 ValProValIleAsnSerMetGlySerSerLeuThrLeuGlnProValGlnPheSer 465
 QY 939 -----ATGTGTCCAAATAGCAATGCTCTTACT 965
 Db 466 GlnProLeuHisProSerTyrGlnGlnProLeuMetProProValGlnSerHisValThr 485
 QY 966 GCAGACCCAGGATGCTCCACCTGTCTCTGGCATACCCACATGCA-----AAGCAC 1019
 Db 486 GlnSerProPheMetAlaThrMetAlaGlnLeuGlnSerProHisAlaLeuTyrSerHis 505
 QY 1020 AAAGAACATTTATCC-----ATACATCTCAATATGGTCCCAAGTGTGCACATGCAC 1073
 Db 506 LysProGluValAlaGlnTyrThrHisThrGlyLeuLeuProGlnThr-----MetLeu 523
 QY 1074 GTAACACACACACACACAAATTCAGGTAGCAGGTACGTGGCAAGTATATTCTGCTCATC 1133
 Db 524 IleThrAspThr----- 527
 QY 1134 AAATGGTCAATTGGCTATGTACTTTTGGCAGGGAAGTACATTATCTACAGTCACAAAATG 1193
 Db 528 -----ThrAsnLeuSerAlaLeuAlaSerLeu 536
 QY 1194 TCTCATGGGAAAGCCTTGCAGATTCAGACACATATACAAATTCCTACACGACAGGC 1253
 Db 537 ThrProThrLysGlnValPheThrSerAspThr-----GluAlaSerSerGluSerGly 554
 QY 1254 CCCCATACCATCTATTCCATAACCACTCAGGTTACAGATGATGCTTCTTCTATTCT 1313
 Db 555 LeuHisThrProAlaSerGlnAlaThrLeuHisValProSerGlnAspProAlaGly 574
 QY 1314 -----AACTCTACACATAAACTTTTACTGGAAGTACTCATATAATTGGACATTC 1361
 Db 575 IleGlnHisLeuGlnProAlaHisArgLeuSerAlaSerProThrValSerSerSerSer 594
 QY 1362 AGCAACTGTCTAGTCTCC 1379

```

Db      595 LeuValLeuTyrGlnSer 600
RESULT 5
JQ0107
hypothetical 66K protein - Ononis yellow mosaic virus
C:Species: Ononis yellow mosaic virus
A:Note: host Pisum sativum cv. Early massey (pea)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: JQ0107
R;Ding, S.W.; Keese, P.; Gibbs, A.
Virology 172, 555-563, 1989
A:Title: Nucleotide sequence of the ononis yellow mosaic tymovirus genome.
A:Reference number: JQ0106; MUID:90021186; PMID:2800337
A:Accession: JQ0107
A:Molecule type: genomic RNA
A:Residues: 1-597 <DIN>
A:Cross-references: GB:J04375; NID:G332572; PIDN:AAA46795.1; PID:G332573

Alignment Scores:
Pred. No.:      0.189      Length:      597
Score:          112.00     Matches:     96
Percent Similarity: 33.25%  Conservative: 45
Best Local Similarity: 22.64% Mismatches:    121
Query Match:     1.40%     Indels:     163
DB:             2         Gaps:       24

US-10-776-827-10 (1-4460) x JQ0107 (1-597)
QY      2656 CCTCTCACCTATATGCGCCATAAAGTGCCTAAGCACTCAGGCCTC----- 2700
Db      187 ProLysProLeuHisLeuHisAsnPheArgGlnHisSerSerLeuCysSerArgArgSer 206
QY      2701 -----CCACTCATCAACCCCTTTGACCAGAGAAAGAACAC 2736
Db      207 ProCysArgLysLeuArgProThrProArgCysAsnGlnLeuAlaGlnGlnHis 226
QY      2737 -----TCGTGTTCTATATCCCTTCTCACAATA 2763
Db      227 ProLeuProSerSerLysProLeuSerLeuGlnAlaGlyLeuGlyProCysProLeu 246
QY      2764 GAGAGTTTGTGATGGGGCCTCTGGCTGTGCGCTTCCATCAACAGATAAATGCGCATCTG 2823
Db      247 -----ProProHisAsnLys----- 251
QY      2824 CTTGCACCAAAACCCAGGATGTGGAGACATCT-----CCCCACAATGCCAC 2871
Db      252 ArgSerProSerProAlaValIleArgLysThrAlaGlyLeuLeuProHisProLysLeu 271
QY      2872 TGTCCACAGACAAGTCGCCCTCTCTGCTCCACCTCTCAGTCCGCCCTAGATGGATGG 2931
Db      272 ProProSerSerArgGlyHisLeuProSerSerThrSerSer----- 286
QY      2932 CTGGGGAGAGTGGAGGCTGACAGCTGAGACGTAGTGTCTCAGATATGATCTAGGAGGCGG 2991
Db      287 -----SerProArgSerAsnArgGlyValArg 295
QY      2992 ATCACCAGGATCCGGGACCATACAAGTAACATGTTTCCATG---GCAACTGTGTC--- 3045
Db      296 CysSer---ValHisLeuHisLysSerArgSerAsnSerGlnAspLeuArgSerCysArg 314
QY      3046 -----TCGTTTGNATTAAGACAGCAGTCAGTTGTCTATTGCCATGACAGCCCTCTA--- 3096
Db      315 ValArgSerAsnSerLeuGlnGlnThrProIleLeuMetGlyHisPheLysSerLeuGly 334
QY      3097 ---TCTCCAGCACAATGTCCTCTCTCTAATCCA-----ATG 3135
Db      335 GlnSerProAsnLeuArgSerPheGluArgProArgProThrArgArgSerLeuArgLeu 354
QY      3136 GACTTGCTCTCACCCAGGAGTGAACACCCAGAACTCCTCTCTCAG----- 3183
Db      355 LeuProLeuSerProGlnLysValProThrValHisValProThrHisGlnGlnSerGly 374
QY      3184 -----TCATCTCCACGCCGATGACTCAGAAGAGCCAAACCCAGATGGGGCC 3231

```

```

Db      375 HisLysGlyProSerLeuProArgProHisSerProSerArgGlnThrHisHisAla--- 393
QY      3232 TCTCTTTTCCCATCACAGACTCCCTGACAACCTTTCTCGCGTAACCTAGAGAGTCC 3291
Db      394 -----ArgLeuPro 396
QY      3292 AGTGCAGATAGGCCCTAAACGTTTGTAAATAAACACAGTG-----CATGAAGAGAGC 3345
Db      397 His-----SerLysArgValSerLeuProAsnSerValLeuHisHisAspArgPro 413
QY      3346 CTAAGGCCATTGTTGATATCCACTCTCTCTCTTCCACTT-----CCTTCTCATCTT 3396
Db      414 LysArgProIleHisPheGlySer-----PheProIleAsnValAlaProSerHisLeu 431
QY      3397 TTTCTCCATGTTTATGCTTCTCTGATCCCTCTTCTGCTGCTGCACAGACCCAGC 3456
Db      432 Leu-----ProArg 434
QY      3457 CCTTATTTCTCTCCATTTTCACTCTCTCCAGCCTCTGTCCTGAAGTCCACT---GGC 3513
Db      435 LysLeuTrpSerArgAlaSerSerProProThrCysSerProThrThrSerAsnHisGly 454
QY      3514 AACCCATGGGAC---CTCAGGACACGAGACTGCTTGACTCATCTGGGAGGAGTAAGTTCA 3570
Db      455 HisProGluGluAlaLeuArg----- 461
QY      3571 CGGGGGACAAAAAATGATTTCTTAAAGAGAGGCTTCTTAGACACACAGAGC----- 3623
Db      462 -----PheLeuProLysAsnLeuProGlnHisCysGlnMetAlaLeu 475
QY      3624 -----TCCAGAAAGACATCCCTTAGCCCTGGACTTCTGACGACCTTTAGC 3668
Db      476 MetGluAsnTyrCysSerHisPheSerSerPro-----SerSerSerVal 490
QY      3669 CAGGCTCCGAGCGGACGAGGAGGCGCTTTCCCATTTGCTCTTCCCATTTGCTCAA 3728
Db      491 SerPheProGluAspHisGlnSerSerLeuProProIleSer-----ThrArg 506
QY      3729 TGGATTCCCATGT 3740
Db      507 TrpValGlnCys 510

RESULT 6
T45746
hypothetical protein F24M12.190 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45746
R;Vitale, D.; Liguori, R.; Flores, M.; Argiriou, A.; De Simone, V.; Mewes, H.W.; Lemcke,
submitted to the Protein Sequence Database, December 1999
A:Reference number: Z23012
A:Accession: T45746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-968 <VIT>
A:Cross-references: EMBL:AL132980
A:Experimental source: cultivar Columbia; BAC clone F24M12
C:Genetics:
A:Map position: 3
A:Introns: 69/1; 99/1; 132/3; 170/3; 198/1; 220/3; 242/3; 284/2; 396/3; 441/3; 514/1; 81
A:Note: F24M12.190

Alignment Scores:
Pred. No.:      0.384      Length:      968
Score:          109.00     Matches:     99
Percent Similarity: 36.65%  Conservative: 74
Best Local Similarity: 20.97% Mismatches:    150
Query Match:     1.36%     Indels:    149
DB:             2         Gaps:       23

US-10-776-827-10 (1-4460) x T45746 (1-968)

```


QY	930	TTGGCAGCTATGTGTCACATAGCAATGCTCCTTACTGACAGCCAGGATGCTCCACC	989
Db	79	-----	79
QY	990	TGTCTCTGGCATACCCACATGCAAGCAAAAGCAATTATTCATCATCTCAATATG	1049
Db	79	-----	79
QY	1050	GTTCACCAAGTGTGCACATGACGTAACTACACACACACACAAATTCAGGTAGCAGGTAC	1109
Db	80	-----	93
QY	1110	GTGGCAAGTATATCTGCTCATCAATGGTCAATGGCTATGTACTTTGTGCAGGAAGT	1169
Db	93	-----	93
QY	1170	ACATTATCTACAGTCACAAAATGTCTCATGGGAAGCCTTGGCAGATTTCAGACACATAT	1229
Db	94	-----	108
QY	1230	ATACATTTCTTACACCAAGGCCCATACACCATCTATTCCATAAACCACTCAGGTT	1289
Db	109	-----	120
QY	1290	ACAGATGCTATCTTCTTATTTCTAACTCTACACATAAACTTTTACTGGAAGTACTATA	1349
Db	121	-----	137
QY	1350	ATTGGACATTCAGCAACTG	1370
Db	138	LeuGlySerGluAlaAsnLeuThrCysThrLeuThrGlyLeuArgAspAlaSerGlyVal	157
QY	1371	-----	1394
Db	158	ThrPheThrTrpThrProSerSerGlyLysSerAlaValGluGlyProProGluArgAsp	177
QY	1395	TTGATACACACACACCAAGTTTCTGTGCTCTGACCCCTCAGCTGTGCCAAGATGTTAA	1454
Db	178	LeuCysGlyCysTyrSerValSerValLeuPro---GlyCysAlaGluProTirpAs	196
QY	1455	AGTGTGATGGTTCAAAATTCATGAAGCTTTTCTTCTTAACTCATGACAAGTCCGTC	1514
Db	196	n-----HisGlyLysThrPheThrCysThrAlaAlaTyrProGluSe	210
QY	1515	CTCATTTGCCACTCAGAGGTGTTTAAATGTATCCA-----AGACCTCTCTGTGA	1562
Db	210	rLysThrProLeuThrAlaThrLeuSerLysSerGlyAsnMetPheArgProGluValHi	230
QY	1563	AACATTACCCCGCAAACTCAG	1587
Db	230	sLeuLeuProProProSerGluGlu	238
RESULT 8			
JQ0110			
hypoetical 69K protein - turnip yellow mosaic virus			
C:Species: turnip yellow mosaic virus, TYMV			
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 19-May-2000			
R:Keese, P.; Mackenzie, A.; Gibbs, A.			
Virology 172, 536-546, 1989			
A:Title: Nucleotide sequence of the genome of an Australian isolate of turnip yellow mosaic virus			
A:Reference number: JQ0109; MUID:90021184; PMID:2800335			
A:Accession: JQ0110			
A:Molecule type: genomic RNA			
A:Residues: 1-628 <KEE>			
A:Cross-references: GB:J04373; NID:g332244; PIDN:AAA46591.1; PID:g332245			
A:Experimental source: Australian isolate			
C:Superfamily: hydroxyproline-rich glycoprotein			
Alignment Scores:	0.55	Length:	628
Pred. No.:	107.00	Matches:	96
Score:			

Percent Similarity:	30.57%	Conservative:	37
Best Local Similarity:	22.07%	Mismatches:	155
Query Match:	1.33%	Indels:	147
DB:	2	Gaps:	17
US-10-776-827-10 (1-4460) x JQ0110 (1-628)			
QY	2748	ATCCCTTGTGCATAGAGAGTTTGTTCATGGGCGCT	2798
Db	40	LeuProMetValHisSerGluGlyThrSerAlaProThrGlnLeuLeuArgHisProAsn	59
QY	2799	ACATAACAGAAATACCTTGCATCTGCCTGCACCAACCCAGGATGGAAGACATCTCC	2858
Db	60	lIetrPheGlyAsnIleProProProArgProGlnAspAsnArgAspPheSer	79
QY	2859	CCAACTCGCAGCTG---CTCACCAGGACAAAGCTGCCCTTCTCTCCACCTCTCAGTC	2915
Db	80	ProLeuHisProLeuValPhe-ProGlyHisHisSer	91
QY	2916	CCCCTAGATGGTGGTGGGAGGCTGACAGCTGAGAGCTAGTGTGCAGATA	2975
Db	92	-----GlnLeuArg	94
QY	2976	TGATCTAGAGGGCGGATCACCGGATCCGGACATACAAAGTAACATGGTTTCCATGGC	3035
Db	95	-----HisValHisGluThrGlnValG1	103
QY	3036	AACGTCTTGTCTGGTTGAATTA	3057
Db	103	nGlnThrCysProGlyGluLeuLysLeuSerGlyIleGluLeuProProAlaProG1	123
QY	3058	AGACAGCAGTCAGTTGTCATTGCCATCAGAGGCTCTATCT---CCAGGCACAAATGTC	3113
Db	123	mArgHisSerLeuProLeuHisIleThrArgProSerArgPheProHisPheHi	143
QY	3114	CCTGTGCTCTCTAATCCAATGGACTTCTCTCACCAGGATGAAACACCCAGAACT	3173
Db	143	salaArgArgProAspValLeuProSerLeuProAspHisGly	159
QY	3174	CACCTTCTCAGTCATCTCCACAG---CCGATGACTCAGAGAGCCAAACCCAGATGGGCG	3230
Db	159	lLeuAlaGluThrLysProArgThrSerValArgGlnProArgSerThrThrArgGlyPr	179
QY	3231	CTCTCTTTTCCCATCACACA-----GACTCCCTCGACAACT	3266
Db	179	oSerPheArgProIleLeuLeuProGluValValHisAspAspProHisSe	199
QY	3267	TTCTCTGGCTAACTAGAGGAGTCCACAGTGCAGGATAGGCCCTTAACGTTTGTAAATAA	3326
Db	199	rSerLeuArgProGlyArgSerArgSer-----Ar	209
QY	3327	ACAGGTGCATGAAAGGAGCCTAAGGCATTTGTATATCCACTCTCTCTTT	3378
Db	209	gGlnLeuGlnProThrIleArgArgProLeuLeuAlaProAsnGlnPheAsnSerProAr	229
QY	3379	-----CCACTTCTCTCTCATCTTTTCTCCATGTTTATGCTTCTCTCTATCCCT	3428
Db	229	gGlnProProProLeuSerAspAspProGlyIleLeuGlyProCys-ProLeuAlaProA	249
QY	3429	CTTCTGCTGCACACAGACCCAGCCCTTTATTCCTCTCTCCATTTT-----CACTCC	3482
Db	249	snSerThrArgAspProPro-ProArgProIleThrProGlyProPheAsnThrHisGly	268
QY	3483	TTCCAGCCTCTGCTCCCTG-----	3500
Db	269	ValArgProLeuSerValLeuProArgAlaProProArgArgGlyLeuLeuProAsnPro	288
QY	3501	-----AACTGCCACTGGCAACCATGGGA	3524
Db	289	ArgArgHisArgThrSerThrGlyHisIleProSerThrThrAlaSerArgProThrGly	308
QY	3525	CCTCAGGAC-----CAGAGACTGCTTGAATCTCTGGGAGGTAAGTTCACGGGGGAC	3578

Db	2160	eIleArgThrThrglyLeuArgProTyrProSerSerValleuIleCysCysValleuAs	2180
Qy	3328	C-----AGGTGCATGAAGAGGACCTAAGGCCAT---TGTTG	3360
Db	2180	nAspThrTyrAlaProGlyGluGluValTyAsnGlyThrTyrGlyAspThrCysTy	2200
Qy	3361	ATATCCACTCTCTCTTCCACTCTCTCTCATCTTTTCTCCATGTTTATGCTTCTCT	3420
Db	2200	rPheValAsnCysSerLeuSerCysThrLeu-----GluPheTyAsnTrpSe	2216
Qy	3421	GATTCCCTCTTCTGCTGCACACAGCCAGCCGCCCTTTATTCCTCTCCATTTCAC	3480
Db	2216	rCysProSerThrProSerProThrProThrProSer---LysSerThrProThrProSe	2235
Qy	3481	CTTCCAGCCCTGCTCCCTGAAGTGCACCTGCACCCACCTCAGGACCCAGAG	3538
Db	2235	rLysProSerSerThrProSerLysPro-----ThrProGlyThrLysPProGlu	2252
RESULT 10			
RBHUAP			
adenomatous polyposis coli protein - human			
A:Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC			
C:Species: Homo sapiens (man)			
C:Date: 31-Dec-1991 #sequence revision 31-Dec-1991 #text_change 21-Jul-2000			
C:Accession: A37261; B39658; A44928; A49319; I54271			
R:Kinzler, K.W.; Nilbert, M.C.; Su, L.X.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith			
chul, S.P.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.			
Science 253, 661-665, 1991			
A:Title: Identification of FAP locus genes from chromosome 5q21.			
A:Reference number: A37261; MUID:91335210; PMID:1651562			
A:Accession: A37261			
A:Molecule type: mRNA			
A:Residues: 1-2843 <KIN>			
A:Cross-references: GB:M74088; NID:g1823396; PIDN:AAA03586.1; PID:g1823397			
R:Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grod			
arrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Le			
Cell 66, 601-613, 1991			
A:Title: Identification of deletion mutations and three new genes at the familial polypos			
A:Reference number: A39658; MUID:91330307; PMID:1678319			
A:Accession: B39658			
A:Molecule type: DNA			
A:Residues: 1-183,'L',185-969,'N',971-1308,'G',1310-1324,'SS',1326,'HSTLE',1332-1354,'P'			
A:Cross-references: GB:M73548; NID:g190163; PIDN:AAA60354.1; PID:g190164			
R:Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst			
Cancer Res. 52, 643-645, 1992			
A:Title: Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a c			
A:Reference number: A44928; MUID:92119623; PMID:1310068			
A:Accession: A44928			
A:Molecule type: DNA			
A:Residues: 1506-1525 <MIK>			
A:Cross-references: GB:S78214; NID:g243541; PIDN:AAB21145.1; PID:g243542			
A:Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBIIP:78218)			
R:Spurio, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber			
Cell 75, 951-957, 1993			
A:Title: Alleles of the APC gene: an attenuated form of familial polyposis.			
A:Reference number: A49319; MUID:94073973; PMID:8252630			
A:Accession: A49319			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 'G',143-171,'P',173-179 <SPI>			
A:Cross-references: GB:S67787; NID:G461061; PIDN:AAD13997.1; PID:g4261697			
R:Lambertz, S.; Ballhausen, W.G.			
Hum. Genet. 90, 650-652, 1993			
A:Title: Identification of an alternative 5' untranslated region of the adenomatous poly			
A:Reference number: I54271; MUID:93186137; PMID:8383094			
A:Accession: I54271			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1-4 <LAM>			
A:Cross-references: GB:S56365; NID:g266243; PIDN:AAD14918.1; PID:g4262770			
A:Gene: GDB:APC			
C:Genetics:			
A:Cross-references: GDB:119682; OMIM:175100			

A:Map position: 5q21-5q22			
A:Note: mutations of this gene can result in familial adenomatous polyposis or sporadic c			
C:Superfamily: adenomatous polyposis coli protein			
C:Keywords: cancer; familial adenomatous polyposis; tumor suppressor			
F:1-730/Domain: leucine-rich <NTD>			
F:7-72/Region: coil #status predicted			
F:185-227/Region: coil #status predicted			
F:731-2832/Domain: serine-rich <CTD>			
F:1131-1156/Region: acidic			
F:1558-1577/Region: acidic			
F:1866-1893/Region: highly charged			
Alignment Scores:			
Pred. No.:	0.766	Length:	2843
Score:	106.50	Matches:	75
Percent Similarity:	35.41%	Conservative:	33
Best Local Similarity:	24.59%	Mismatches:	105
Query Match:	1.33%	Indels:	92
DB:	1	Gaps:	14
US-10-776-827-10 (1-4460) x RBHUAP (1-2843)			
Qy	2797	TCACATAACAGAAATAACTTGCATCTGCCTGCACCAACCAGGATGTGGAGACATCT	2856
Db	2225	SerArgGlyArgThrMetIleHisIleProGlyValArgAsnSerSerSerThrSer	2244
Qy	2857	CCCACAACTGCCACTGCTCACCAGGACAGCTGCCCTTCTCTCCACCTCTCAGTCC	2916
Db	2245	ProValSerLysGlyProPro-----LeuLysThrProAlaSerLysSer	2260
Qy	2917	CCCTAGATGATGCTGGGAGAGCTGAGAGCTGACAGCTGAGAGCTAGTCTCAGATAT	2976
Db	2261	ProSerGluGlyGlnThrAlaThrThrSerProArgGlyAlaLysProSerValLysSer	2280
Qy	2977	GATCTA-----GGAGGCGGATC-----	2994
Db	2281	GluLeuSerProValalaArgGlnThrSerGlnIleGlyGlySerSerLysAlaProSer	2300
Qy	2995	---ACGGGATCCGGACCATACATACTGTTTCCATGGCAACTGCTTCTGCTGTTT	3051
Db	2301	ArgSerGlySerArgAspSerThrProSerArgProAla-----	2313
Qy	3052	GAATTAAGACAGACAGTCACTGTTCATTGCATGCATGACAGGCT-----CTATCT	3099
Db	2314	-----GlnGlnProLeuSerArgProIleGlnSerProGlyArgAsnSerIleSer	2330
Qy	3100	CCAGGCACAAATGCTCCTGCTGCTCTCTAATCCAAATGAGCTGCTC-----TCA	3147
Db	2331	ProGlyArgAsnGlyIleSerProProAsnLysLeuSerGlnLeuProArgThrSerSer	2350
Qy	3148	CCCAGGGATGAACACACCCAGAACTCACTTCT-----	3180
Db	2351	ProSerThrAlaSerThrLysSerSerGlySerGlySerGlySerMetSerTyrThrSerProGly	2370
Qy	3181	---CAGTCACTCCACAGCCGATGATGCTCAGAGAGCCAAACCCAGGAATGGGGCTCTCTT	3237
Db	2371	ArgGlnMetSerGlnGlnAsnLeuThrLysGlnThrGlyLeuSerLysAsnAlaSerSer	2390
Qy	3238	TTCCCATCATCAGACTCCCTGACAACTTTCCTGGGTAACCTAGAGGAGTCCACAGTCA	3297
Db	2391	IlePro-----ArgSerGluSerAla	2397
Qy	3298	GGATAGGCCCTAAACGTTTGT-----TTAAATAACACAGGTGCAT-----	3336
Db	2398	SerLysGlyLeuAsnGlnMetAsnAsnGlyAsnGlyAlaAsnLysLysValGluLeuSer	2417
Qy	3337	-----GAAAGGAGCCCTAAGGCCATTTGT	3360
Db	2418	ArgMetSerSerThrLysSerSerGlySerGluSerAspArgSerGluArgProValLeu	2437
Qy	3361	ATA-----TCCACTCTCTTCTTCCACTTCTCTTCTCATCTTTTTC-----	3400
Db	2438	ValArgGlnSerThrPheIleLysGluAlaProSerProThr-LeuArgArgLysLeuG1	2457


```
QY 3145 TCACCCAGGATGAACACCCAGAAACTCAGTCTCTCAGTCACTTCCA----- 3192
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
139 AlaProCysSer---AlaSerSerGluLeuSerGlyProThrProLeuHisThrSer 157
QY 3193 CAGCCGATGACTCAGAGAGCCAAACCCAGAAATGGGCGCTCTCTTTCCCGATCAGAC 3252
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
158 SerProValGlnGlyLysGluSerThrArgGlnAspThrArgSerProThrSerAla 177
QY 3253 TCCCTGCACAACTTCTCTGGCGTAAGTCTAGAGGAGTCCC-----AGTGCAGGATAGGCC 3306
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 ProThrGlnArgLeuSerProGlyGluAlaLeuProValSerGlnAlaGlyThrGly 197
QY 3307 CTAACCGTTTGTAAATAAACAGGTGCATGAAGAGCCCTAAGGCCCATTTGTATATCC 3366
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
198 LysAlaProGluLeuProSerGlySerGlyProGluProLeuAlaProSerProValSer 217
QY 3367 ACTCTCTCTTCCACTTCTCTCATCTTTCTCTCATGTTTATGCTTCTCTGATTC 3426
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
218 ProThrPheProPro-----SerSer-Pr 225
QY 3427 CTCTCTGCTGCTGCACACAGCCAGCCCGCCCTTTATTC-----TCTCCATTTTC 3477
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
225 OserAspTyrProGlnGluArgSerProGlyGlyHisSerAspGlyAlaSerProArgSe 245
QY 3478 ACTCCTTCCAGCCTCTCTCCCTGAATG-----CCACTGGCAA---CCCATGGGAC 3525
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
245 rProValProThrThrLeuProGlyLeuArgHisAlaPro-TyrGlnGlyProArgGlyp 265
QY 3526 CTCAGGACACAGACTCTGTGACTCTCTGGGAGGGTAAGTTCACGGGGGACAAAAAA 3585
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
265 roProAsp----- 267
QY 3586 TGATTCCTAAAGAGAGGCTTCTAGACACAGCAGGCTCCAGAAAGACA-----T 3636
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
268 -----SerProaspGlySerProLeuThrProValLeuS 279
QY 3637 CCCTAGGCTGGACTTCTGAGCAGCTTTAGCCAGGCTCCGACGGCGGACGAGGAGGC 3696
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
279 erGlnMetProTyrLeuValAlaSer---ProGluProProGlnSerSerProThrProA 298
QY 3697 CTTTCCCATGCT-----CCTTCCCATTTG 3723
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
298 laPheProLeuAlaAlaSerTyrAspThrAsnGlyLeuSerGlnProProLeuProGluL 318
QY 3724 CTCATCGATTCCATGTTCTTTTCTTGGGGGAGCAGGAGGAGAAAGGT 3776
Db   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
318 ysArgHisLeuPro-----GlyProGlyGlnGlnProGly 329
```

Search completed: September 21, 2004, 15:44:56
Job time : 197 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: September 21, 2004, 15:03:24 ; Search time 374.5 Seconds
(without alignments)
6729.836 Million cell updates/sec

Title: US-10-776-827-10
Perfect score: 8026
Sequence: 1 cggggcagcaaccaggagat.....aattaaagggtactactagctc 4460

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO spool_p/US10776827/runat 21092004 133239 20757/app query.fasta 1.4615
-DB=A_Geneseq 26Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10776827@cgn 1.1.651 @runat 21092004 133239 20757 -NCFU=6 -ICFU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 29Jan04.*

1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	507.5	6.3	99	6 ABO00556	Novel hum
2	267	3.3	47	5 AAM49810	Human KIA
3	249.5	3.1	49	5 AAM49809	Human KIA
4	170	2.1	88	4 AAG75725	Human col
5	159	2.0	28	4 AAO11854	Human pol
6	159	2.0	28	6 ABO00836	Polypepti
7	127.5	1.6	1213	6 ABU11838	Human MDD
8	127	1.6	390	4 AAU08755	Human ins
9	127	1.6	390	4 ABG01683	Novel hum
10	125.5	1.6	572	5 AAG80596	Human cad

11	120	1.5	899	4 ABG25235	Novel hum
12	118	1.5	487	6 ABM64771	Protonib
13	118	1.5	654	4 ABG08388	Novel hum
14	118	1.5	1078	6 ABP96069	Human pro
15	118	1.5	1251	6 ABU54581	Human NOV
16	114.5	1.4	806	4 ABG04954	Novel hum
17	113.5	1.4	443	4 AAB68080	Amino aci
18	113.5	1.4	451	3 AAY99382	Human PRO
19	113.5	1.4	451	4 AAB66131	Protein o
20	113.5	1.4	451	4 AAU29164	Human PRO
21	113.5	1.4	451	4 AAB87566	Human PRO
22	113.5	1.4	451	4 AAU02749	Novel Hum
23	113.5	1.4	451	4 AAB48069	Human ext
24	113.5	1.4	451	5 ABG95891	Human sec
25	113.5	1.4	451	5 AAU78174	Human cho
26	113.5	1.4	451	6 ABU58540	Human PRO
27	113.5	1.4	451	6 ABU88088	Novel hum
28	113.5	1.4	451	6 ABU84403	Human sec
29	113.5	1.4	451	6 ABR66277	Human sec
30	113.5	1.4	451	6 ABR65667	Human sec
31	113.5	1.4	451	6 ABU99607	Human sec
32	113.5	1.4	451	6 ABU82846	Human PRO
33	113.5	1.4	451	6 ABU89967	Novel hum
34	113.5	1.4	451	6 ABR68216	Human sec
35	113.5	1.4	451	6 ABU96269	Novel hum
36	113.5	1.4	451	6 ABU92700	Human sec
37	113.5	1.4	451	6 ABO08777	Human sec
38	113.5	1.4	451	6 ABO02829	Human sec
39	113.5	1.4	451	6 ABR74983	Human sec
40	113.5	1.4	451	6 ABR94745	Human sec
41	113.5	1.4	451	6 ABU85718	Human PRO
42	113.5	1.4	451	6 ABU98878	Novel hum
43	113.5	1.4	451	6 ABU98093	Novel hum
44	113.5	1.4	451	6 ABU91799	Novel hum
45	113.5	1.4	451	6 ABU89492	Human PRO

ALIGNMENTS

RESULT 1

ABO00556

ID ABO00556 standard; protein; 99 AA.

AC ABO00556;

DT 06-AUG-2003 (first entry)

DE Novel human polypeptide #143.

Human; angiogenesis; cytokine; cell proliferation; pluripotent;
cell differentiation; totipotent; stem cell; transplantation; bio-sensor;
neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;
nerve; brain tissue; central nervous system disease;
peripheral nervous system disease; neuropathy; haematopoiesis; bone;
myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;
regeneration; cartilage; tendon; ligament; nerve tissue growth;
tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;
osteoarthritis; bone degenerative disorder; periodontal disease;
gut protection; lung fibrosis; liver fibrosis; reperfusion injury;
immune deficiency; infection; autoimmune disorder; allergic reaction;
thrombolytic; thrombosis; coagulation disorder; hereditary disorder;
biohythm; circadian cycle; fertility; metabolism; catabolism; anabolism;
nootropic; neuroprotective; antiparkinsonian; anticonvulsant;
haemostatic; vulnery; antiulcer; osteopathic; antiarthritic;
vasotropic; immunostimulant; antibacterial; fungicide; immunosuppressive;
antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide.

Homo sapiens.

WO2003023013-A2.

20-MAR-2003.

CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 88 AA;

Alignment Scores:
Pred. No.: 3.72e-08 Length: 88
Score: 170.00 Matches: 36
Percent Similarity: 97.37% Conservative: 1
Best Local Similarity: 94.74% Mismatches: 1
Query Match: 2.12% Indels: 1
DB: 4 Gaps: 0

US-10-776-827-10 (1-4460) x AAG75725 (1-88)

QY 2844 GTGGAAGACATCTCCCAACACTGCCACTGCTCCACGAGACAGCTGCCCTTCTCTCTC 2903
Db 51 MetGluAspIleSerProGlnLeuProLeuLeuThrArgThrSerCysProSerCysLeu 70
QY 2904 CACCTCTCAGTCCCTCCAGATGGATGGCTGG-GGAGAGGTGGAGGCTGACAGC 2956
Db 71 HisLeuSerValProLeuGluTrpMetAlaGlyGlyGluValGluAlaAspSer 88

RESULT 5

AAO11854
ID AAO11854 standard; protein; 28 AA.

AC AAO11854;

XX
DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 25746.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

XX WO200164835-A2.

XX
PD 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US004927.

XX 28-FEB-2000; 2000US-00515126.

PR 18-MAY-2000; 2000US-00577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI91785.

XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.

XX Claim 20; SEQ ID NO 25746; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 28 AA;

Alignment Scores:
Pred. No.: 2.68e-07 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 4 Gaps: 0

US-10-776-827-10 (1-4460) x AAO11854 (1-28)

QY 783 CCACACACAGACATTTTCAGAACCCCTGAAATAATGACATATGTCATGTCACAGATGA 842
Db 1 ProHisThrAspIleSerGlyThrProGluIleMetHisTyValHisValHisArgVal 20

QY 843 ACTACTCAACCCAGGACAACT 866

Db 21 ThrThrGlnProArgAsnLysPro 28

RESULT 6

ABO00836

ID ABO00836 standard; protein; 28 AA.

XX ABO00836;

XX
DT 06-AUG-2003 (first entry)

XX Polypeptide encoded by novel human contig #87.

XX Human; angiogenesis; cytokine; cell proliferation; pluripotent;
XX cell differentiation; totipotent; stem cell; transplantation; bio-sensor;
XX neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;
XX nerve; brain tissue; central nervous system disease;
XX peripheral nervous system disease; neuropathy; haematopoiesis; bone;
XX myeloid disorder; lymphoid cell disorder; ligament; nerve tissue growth;
XX regeneration; cartilage; tendon; ulcer; osteoporosis; cancer;
XX tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;
XX osteoarthritis; bone degenerative disorder; periodontal disease;
XX gut protection; lung fibrosis; liver fibrosis; reperfusion injury;
XX immune deficiency; infection; autoimmune disorder; allergic reaction;
XX thrombolytic; thrombosis; coagulation disorder; hereditary disorder;
XX biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;
XX neurotropic; neuroprotective; antiparkinsonian; anticonvulsant;
XX haemostatic; vulnery; antidiabetic; osteopathic; antiarthritic;
XX vasotrophic; immunostimulant; antibacterial; fungicide; immunosuppressive;
XX antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide.

XX Homo sapiens.

XX WO2003023013-A2.

XX 20-MAR-2003.

XX 13-SEP-2002; 2002WO-US029001.

XX 13-SEP-2001; 2001US-0322511P.

PR 12-SEP-2002; 2002US-00243552.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Yang Y, Wang Z, Weng G, Ma Y;

XX WPI; 2003-313249/30.

DR N-PSDB; ACD05913.

XX Novel nucleic acids and polypeptides for diagnosis, treatment of central
XX and peripheral nervous system diseases and neuropathies, such as

PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
PT lateral sclerosis.
XX
PS
PS Example 3; SEQ ID NO 960; 300pp; English.
XX
XX The present invention relates to the isolation of novel human
CC polynucleotide sequences and their encoding polypeptides. The novel
CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell
CC proliferation, cell differentiation, anti-inflammatory, and stem cell
CC growth factor activities. The polypeptides are involved in the
CC proliferation, differentiation and survival of pluripotent and totipotent
CC stem cells, and are useful for re-engineering damaged or diseased
CC tissues, transplantation, manufacture of bio-pharmaceuticals and
CC development of bio-sensors. The polypeptides can be used to manipulate
CC stem cells in culture to give rise to neuroepithelial cells that can be
CC used to augment or replace cells damaged by illness, autoimmune disease,
CC accidental damage or genetic disorders. The polypeptides induce the
CC proliferation of neural cells and regeneration of nerve and brain tissue
CC and are useful for the treatment of central and peripheral nervous system
CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,
CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The
CC polypeptides are also involved in chemotactic or chemokinetic activity,
CC regulation of haematopoiesis and are useful for treating myeloid or
CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and
CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, in tissue repair, healing of burns, incisions, ulcers, for
CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and
CC periodontal disease. The polypeptides are also useful for gut protection
CC or regeneration and treatment of lung or liver fibrosis, reperfusion
CC injury in various tissues, various immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid
CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and
CC conditions, such as asthma or other respiratory problems. The
CC polypeptides are involved in thrombolysis or thrombosis and are useful in
CC treatment of various coagulation disorders (including hereditary
CC disorders such as haemophilia) or to enhance coagulation and other
CC haemostatic events in treating wounds resulting from trauma, surgery or
CC other causes. The polypeptides exhibit immune stimulating or immune
CC suppressing activity, and are useful for treating autoimmune diseases or
CC cancer. They also inhibit the growth, infection or function of infectious
CC agents such as bacteria, fungi, viruses, effect biorythms or circadian
CC cycles of rhythms, fertility of male or female subjects, metabolism,
CC catabolism, and anabolism. ABO00750-ABO00950 represent polypeptides
CC encoded by novel contigs assembled in the examples of the present
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 28 AA;

Alignment Scores:
Pred. No.: 2,68e-07 Length: 28
Score: 159.00 Matches: 28
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.98% Indels: 0
DB: 6 Gaps: 0

US-10-776-827-10 (1-4460) x ABO00836 (1-28)

Qy 783 CCACACACAGACATTTTCAGAACCCCTGAAATATGACATATGTCATGTCACAGAGTA 842
|||||
Db 1 ProHisThrAspIleSerGlyThrProGluIleMetHisTyrValHisValHisArgVal 20
|||||

Qy 843 ACTACTCAACACGACGACAACT 866
|||||

Db 21 ThrThrGlnProArgAenLysPro 28
|||||

RESULT 7
ABU11838
ID ABU11838 standard; protein; 1213 AA.
XX

AC ABU11838;
XX
DT 13-FEB-2003 (first entry)
XX
XX Human MDDT polypeptide SEQ ID 785.
DE
XX MDDT: human; disease detection and treatment molecule polypeptide;
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;
KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;
KW gene therapy; protein replacement therapy; cell proliferative disorder;
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;
KW psoriasis; hepatitis.
XX
OS Homo sapiens.
XX
XX WO200279449-A2.
PN
XX 10-OCT-2002.
PD
XX 27-MAR-2002; 2002WO-US009944.
PF
XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-030001P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Daffo GE, Hillman JU, Yu JY, Tuason O, Yap PE, Anshey SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
PI Feralta CH, David MH, Lewis SA, Chen Au, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI: 2003-058431/05.
DR N-PSDB; ABX34828.
DR
XX New purified disease detection and treatment molecule proteins and
PT polynucleotides, useful for diagnosing, treating or preventing cancers
PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
PT or hepatitis.
XX
XX Claim 27; SEQ ID NO 785; 339pp + Sequence Listing; English.
PS
XX This invention describes a novel disease detection and treatment molecule
CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,
CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,
CC antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides
CC and the polypeptides of the invention can be used for gene therapy, of
CC protein replacement therapy and are useful for treating a variety of
CC diseases or conditions. These polypeptides or polynucleotides are
CC particularly useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. cancers including adenocarcinoma,
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), Crohn's
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or
CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded
CC by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1213 AA;

Alignment Scores:
Pred. No.: 0.00478 Length: 1213

[illegible]

RESULT 10
AAG80596
ID AAG8

[illegible]

Db 975 sPheProLeuGluPro-ProPheProGlnValProSerSerPro----SerLeuLeuProH 994
QY 3489 CCTCTGCTCCTGAACCTGCCA----- 3508
Db 994 isValIleProAlaProHisSerProProPheLeuProArgSerProGlnIleProL 1014
QY 3508 ----- 3508
Db 1014 euHisThrProProAlaLeuHisPheSerHisProProAlaHisProSerPheArgSerH 1034
QY 3509 -----CTGGCAACCCATGGGACCTCAGGACACAGACTGCTGCAC---TCATCTGGGGAG 3560
Db 1034 isSerLeuSerValProGlyValLeuSerPro-ArgLeuLeuHisLeuArgSerLeuVal 1053
QY 3561 GGTAAGTTC-----ACGGGGGACAAA-----AAATGATTCTAAAGAAAGAGGCT 3605
Db 1054 GlyArgPheGlnValThrSerSerLysGluProAlaGluProLeuProGlnProThr 1073
QY 3606 TCCTAGACACAGCAGCTCCAGAAAGACATCCCTAGGCTGGCTGGACTTCTGACAGCTTT 3665
Db 1074 SerProThrLeuSerGlySerProLysProSer---ThrProGlnLeuThrSerGluSer 1092
QY 3666 AGCCAGGCTCCGGACGACGAGGAGGCTTCCCATTCCTCTCTCCCATTTGCT 3725
Db 1093 SerAspThrGluAsp-SerAlaGlyGlyProGluArg----- 1105
QY 3726 CAATGGATTCCATGTTTCTTTTCTTGGGGGAGCAGGAGGAGAAAGTAGAAAAATG 3785
Db 1106 -----HisLeuSer-----GluValGluThrLe 1113
QY 3786 GCAGCCACCTTTCAGAAAAAATATAAGGT----- 3817
Db 1113 uGlnThr-LeuGlnLysGluLeuGluAspLeuTyrSerArgLeuGlyLysGlnProp 1133
QY 3818 --CCAAGCTCTATAGTATTGTTCAGTATTTTCTGTAAATTCGAACACACACAAAAAG 3875
Db 1133 roProGlyIleValAlaProAlaAlaMetLeuSerSerArgGlnArgArgLeuSerLysG 1153
QY 3876 AAAAAATTATTATAAATAACTTTGAAAAATGAAAAAGTCTTGATGTAGTCAGATGGTTA 3935
Db 1153 lySerPhe----- 1155
QY 3936 CTTTCTTAACATTAGGTATTATCCCCCACTCAGACATCCTCAGAAATGAT---CAATGCA 3992
Db 1156 -----ProThrSerArgArgAsnSerLeuGlnArgS 1166
QY 3993 GGGACTCTTTCTGTGACACAAATGTCCAGCCCTCCCT-----G 4031
Db 1166 erGluProTrpCys-----ValProAlaIleProAlaThrGlnGluAlaGluAlaG 1183
QY 4032 GTCAACGCTTCGCCATGGTAGAGTCGTAGTCTGAGGATGAGGAATGGCTGCTCAC 4091
Db 1183 lyGluPheLeuGluProGlyArg-----ArgArgLeuGlnAsnHisSerProSera 1200
QY 4092 CCTTG 4096
Db 1200 laLeu 1201

Search completed: September 21, 2004, 15:23:55
Job time : 423.5 secs

This Page Blank (uspto)